



*Учебники НГТУ*  
Серия основана  
в 2001 году

*Ответственные редакторы серии:*

А.С. Востриков  
Н.В. Пустовой

*Редакционная коллегия:*

Ю.А. Афанасьев  
А.Г. Вострецов  
В.В. Губарев  
В.А. Гридчин  
В.И. Денисов  
К.Т. Джурабаев  
В.И. Игнатьев  
К.П. Кадомская  
В.В. Крюков  
Г.Е. Невская  
В.В. Покасов  
Х.М. Рахимьянов  
Ю.Г. Соловсейчик  
А.А. Спектор  
А.И. Шалин  
А.Ф. Шевченко  
Г.М. Шумский

Допущено Министерством Российской  
Федерации в качестве учебного пособия  
для студентов высших учебных  
заведений, обучающихся  
по направлению 552500-Радиотехника  
и специальности 201400-Аудиовизуальная  
техника



# ЦИФРОВАЯ ОБРАБОТКА ИЗОБРАЖЕНИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

УЧЕБНИК



НОВОСИБИРСК

НГТУ

2002

УДК 681.3.01 (075.8)  
Ц 752

***Федеральная программа книгоиздания России***

Коллектив авторов:

*И.С. Грузман, В.С. Киричук, В.П. Косых,  
Г.И. Перетягин, А.А. Спектор*

Рецензенты:

д-р техн. наук, проф. *Г.В. Мамчев*,  
д-р физ.-мат. наук, проф. *В.П. Голубятников*

§

Ц 752      **Цифровая обработка изображений** в информационных системах: Учеб. пособие / И.С. Грузман, В.С. Киричук и др. – Новосибирск: Изд-во НГТУ, 2002. – 352 с.

ISBN 5-7782-0330-6

Цифровая обработка изображений – интенсивно развивающаяся научная область, которая находит все более широкое применение в различных информационных технических системах: радиолокационных, связи, телевизионных и др. В настоящем учебном пособии рассмотрены следующие задачи обработки изображений: дискретизации и квантования, улучшения визуального качества путем поэлементного преобразования, подавления помех и восстановления, геометрического преобразования и привязки, фотограмметрии и стереовидения, выделения контуров и сегментации, распознавания объектов и математической морфологии, кодирования и сжатия, вычислительной томографии.

Учебное пособие предназначено для студентов старших курсов и магистрантов, изучающих цифровые методы обработки изображений.

ISBN 5-7782-0330-6

УДК 681.3.01 (075.8)

© Новосибирский государственный  
технический университет, 2002 г.



Грузман Игорь Семенович —  
доктор технических наук,  
профессор



Киричук Валерий Сергеевич —  
доктор технических наук,  
старший научный сотрудник



Косых Валерий Петрович —  
кандидат технических наук,  
доцент



Перетягин Георгий Иванович —  
кандидат технических наук



Спектор Адександр Аншелевич —  
доктор технических наук,  
профессор

**Лаборатория цифровой обработки сигналов и изображений  
кафедры Теоретические основы радиотехники  
Новосибирского государственного технического университета**

В лаборатории в течение ряда лет проводятся исследования в области статистических методов и алгоритмов компьютерной обработки изображений. Разрабатываемые методы направлены на решение разнообразных задач цифровой обработки информации, представленной в виде изображений.

Основным достижением лаборатории в области фильтрации изображений является возможность создания быстродействующих рекуррентных и параллельно-рекуррентных алгоритмов, осуществляющих многомерную фильтрацию в реальном времени. Решены такие прикладные задачи, как восстановление изображений при нелинейных искажениях, автоматическое обнаружение объектов при гарантированной стабильности правильного обнаружения. Развита новые подходы к задачам совмещения изображений, полученных в разное время и с помощью различных информационных систем.

Полученные алгоритмы используются в создаваемых и модернизируемых информационных системах в различных областях науки и техники, в частности, в геоинформационных системах и аэрокосмических исследованиях, системах автономной навигации и целеуказания и т. п.

В лаборатории ведутся работы по реализации алгоритмов в вычислительной системе на базе микропроцессора NM 6403, позволяющего создавать высокопроизводительные параллельные комплексы различной конфигурации.

Лаборатория готова оказать помощь в решении задач, связанных с цифровой обработкой изображений, и консультировать по вопросам использования процессора NM 6403 в информационных системах.

Наш адрес: 630092, Новосибирск, пр. К. Маркса, 20, НГТУ  
зав. каф. ТОР проф. А.А. Спектор,  
проф. каф. ТОР И.С. Грузман

Тел. 46-08-55, 46-15-37

e-mail; ktor@ref.nstu.ru

## ОГЛАВЛЕНИЕ

|                                                                                                |     |
|------------------------------------------------------------------------------------------------|-----|
| ВВЕДЕНИЕ .....                                                                                 | 10  |
| Глава 1. ДИСКРЕТИЗАЦИЯ И КВАНТОВАНИЕ<br>НЕПРЕРЫВНЫХ ИЗОБРАЖЕНИЙ.....                           | 11  |
| 1.1. Дискретизация непрерывных изображений .....                                               | 12  |
| 1.2. Квантование изображений .....                                                             | 20  |
| Глава 2. УЛУЧШЕНИЕ ВИЗУАЛЬНОГО КАЧЕСТВА ИЗОБРАЖЕНИЙ<br>ПУТЕМ ПОЭЛЕМЕНТНОГО ПРЕОБРАЗОВАНИЯ..... | 30  |
| 2.1. Линейное контрастирование изображения .....                                               | 31  |
| 2.2. Соляризация изображения .....                                                             | 32  |
| 2.3. Препарирование изображения .....                                                          | 34  |
| 2.4. Преобразование гистограмм, эквализация .....                                              | 39  |
| 2.5. Применение табличного метода при поэлементных преобразо-<br>ваниях изображений .....      | 44  |
| Глава 3. ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ .....                                                          | 48  |
| 3.1. Оптимальная линейная фильтрация. Уравнение Винера-Хопфа.....                              | 49  |
| 3.2. Масочная фильтрация изображений при наличии аддитивного<br>белого шума .....              | 54  |
| 3.3. Рекуррентная каузальная фильтрация изображений.....                                       | 58  |
| 3.4. Применение фильтра Винера для некаузальной двумерной<br>фильтрации .....                  | 62  |
| 3.4.1. Двумерное дискретное преобразование Фурье .....                                         | 64  |
| 3.4.2. Циклическая свертка.....                                                                | 68  |
| 3.4.3. Решение уравнения Винера-Хопфа в циклическом<br>приближении .....                       | 71  |
| 3.5. Байесовская фильтрация изображений.....                                                   | 75  |
| 3.5.1. Сущность байесовской фильтрации .....                                                   | 76  |
| 3.5.2. Марковская фильтрация одномерных последовательностей.....                               | 78  |
| 3.5.3. Двухэтапная марковская фильтрация изображений .....                                     | 81  |
| 3.6. Медианная фильтрация .....                                                                | 86  |
| Глава 4. ВОССТАНОВЛЕНИЕ ИЗОБРАЖЕНИЙ .....                                                      | 93  |
| 4.1. Модели изображений и их линейных искажений.....                                           | 94  |
| 4.1.1. Формирование изображений.....                                                           | 94  |
| 4.1.2. Размытие вследствие движения (смаз) .....                                               | 97  |
| 4.1.3. Расфокусировка .....                                                                    | 99  |
| 4.2. Алгебраические методы восстановления изображений .....                                    | 105 |
| 4.3. Методы восстановления изображений на основе пространствен-<br>ной фильтрации .....        | 107 |

|                                                                                                             |            |
|-------------------------------------------------------------------------------------------------------------|------------|
| 4.3.1. Инверсный фильтр.....                                                                                | 108        |
| 4.3.2. Фильтр Винера.....                                                                                   | 112        |
| 4.3.3. Компенсация краевых эффектов при восстановлении<br>линейно искаженных изображений.....               | 116        |
| 4.4. Итерационные методы восстановления изображений.....                                                    | 122        |
| <b>Глава 5. ГЕОМЕТРИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ И ПРИВЯЗКА<br/>    ИЗОБРАЖЕНИЙ.....</b>                           | <b>130</b> |
| 5.1. Геометрические преобразования на плоскости и в пространстве.....                                       | 131        |
| 5.1.1. Точки и прямые линии на плоскости — двойственность<br>описаний.....                                  | 132        |
| 5.1.2. Однородные координаты.....                                                                           | 134        |
| 5.1.3. Евклидовы преобразования.....                                                                        | 135        |
| 5.1.4. Аффинные преобразования.....                                                                         | 136        |
| 5.1.5. Проективные преобразования.....                                                                      | 140        |
| 5.1.6. Полиномиальное преобразование.....                                                                   | 148        |
| 5.1.7. Оценивание параметров преобразования.....                                                            | 148        |
| 5.2. Восстановление изображения в преобразованных координатах.....                                          | 151        |
| 5.3. Привязка изображений.....                                                                              | 157        |
| 5.3.1. Корреляционный критерий сходства.....                                                                | 157        |
| 5.3.2. Локальное уточнение сдвига.....                                                                      | 161        |
| 5.3.3. Кросс-спектральная мера сходства.....                                                                | 162        |
| 5.3.4. Привязка по локальным неоднородностям.....                                                           | 165        |
| <b>Глава 6. ФОТОГРАММЕТРИЯ И СТЕРЕОВИДЕНИЕ.....</b>                                                         | <b>168</b> |
| 6.1. Модель регистрирующей камеры.....                                                                      | 169        |
| 6.2. Связь между различными системами координат.....                                                        | 171        |
| 6.3. Стереоскопическая система.....                                                                         | 172        |
| 6.4. Калибровка камеры.....                                                                                 | 176        |
| 6.5. Взаимное ориентирование.....                                                                           | 182        |
| 6.6. Поиск сопряженных точек.....                                                                           | 185        |
| <b>Глава 7. СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ.....</b>                                                                | <b>194</b> |
| 7.1. Сегментация изображений на основе пороговой обработки.....                                             | 195        |
| 7.2. Сегментация изображений на основе марковской<br>фильтрации.....                                        | 204        |
| 7.3. Байесовская сегментация, основанная на распределении<br>Гиббса.....                                    | 210        |
| 7.3.1. Распределение Гиббса и его применение к описанию<br>случайных дискретных сигналов и изображений..... | 211        |
| 7.3.2. Байесовская сегментация изображения на основе<br>стохастической релаксации.....                      | 215        |
| <b>Глава 8. ВЫДЕЛЕНИЕ КОНТУРНЫХ ЛИНИЙ.....</b>                                                              | <b>222</b> |
| 8.1. Градиентные методы подчеркивания контуров.....                                                         | 223        |
| 8.2. Ранговое обнаружение локальных контурных признаков.....                                                | 231        |
| 8.2.1. Декорреляция фона изображения.....                                                                   | 234        |
| 8.2.2. Ранговое обнаружение ЛКП.....                                                                        | 238        |
| <b>Глава 9. РАСПОЗНАВАНИЕ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ.....</b>                                                 | <b>246</b> |
| 9.1. Основные положения.....                                                                                | 246        |
| 9.2. Разделяющие функции.....                                                                               | 250        |

|                                                                        |            |
|------------------------------------------------------------------------|------------|
| 9.3. Вычисление информативных признаков.....                           | 253        |
| 9.4. Алгоритмы классификации (краткий обзор) .....                     | 255        |
| 9.5. Байесовский метод распознавания.....                              | 257        |
| 9.6. Вероятности ошибок классификации.....                             | 257        |
| 9.7. Проблема <i>K</i> - классов .....                                 | 267        |
| <b>Глава 10. МАТЕМАТИЧЕСКАЯ МОРФОЛОГИЯ И ОБРАБОТКА</b>                 | <b>270</b> |
| <b>ИЗОБРАЖЕНИЙ</b> .....                                               | <b>273</b> |
| 10.1. Операции математической морфологии .....                         | 275        |
| 10.2. Морфологические операции в дискретном пространстве .....         | 284        |
| <b>Глава 11. СЖАТИЕ ИЗОБРАЖЕНИЙ</b> .....                              | <b>292</b> |
| 11.1. Алгоритмы сжатия без потерь.....                                 | 293        |
| 11.2. Свойства цифровых изображений .....                              | 294        |
| 11.3. Дискретные преобразования изображений в сжатии данных.....       | 296        |
| 11.4. Методы сжатия на основе предсказания .....                       | 301        |
| 11.5. Другие алгоритмы .....                                           | 303        |
| <b>Глава 12. КОМПЬЮТЕРНАЯ ТОМОГРАФИЯ</b> .....                         | <b>307</b> |
| 12.1. Получение проекций.....                                          | 308        |
| 12.2. Классическая томография .....                                    | 315        |
| 12.3. Алгоритм обратного проецирования.....                            | 316        |
| 12.4. Теорема о центральном сечении .....                              | 321        |
| 12.5. Фурье-алгоритм восстановления томограмм.....                     | 323        |
| 12.6. Восстановление томограмм с помощью обратного проецирования ..... | 329        |
| 12.7. Итерационные алгоритмы восстановления томограмм .....            | 338        |
| 12.8. Флюктуационные искажения проекционных данных .....               | 342        |
| Заключение .....                                                       | 345        |
| Литература .....                                                       | 346        |

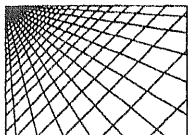
## ВВЕДЕНИЕ

Многие отрасли техники, имеющие отношение к получению, обработке, хранению и передаче информации, в значительной степени ориентируются в настоящее время на развитие систем, в которых информация имеет характер изображений. Изображение, которое можно рассматривать как двумерный сигнал, является значительно более емким носителем информации, чем обычный одномерный (временной) сигнал. Вместе с тем решение научных и инженерных задач при работе с визуальными данными требует особых усилий, опирающихся на знание специфических методов, поскольку традиционная идеология одномерных сигналов и систем мало пригодна в этих случаях. В особой мере это проявляется при создании новых типов информационных систем, решающих такие проблемы, которые до сих пор в науке и технике не встречались и которые решаются сейчас благодаря использованию информации визуального характера.

В связи с этим в вузовских программах появляются дисциплины, направленные на изучение принципов обработки изображений, причем приоритетное внимание уделяется цифровым методам, привлекательным своей гибкостью. Отсутствие учебной литературы является сильным препятствием данному изучению, что и побудило авторов к написанию пособия. Авторы пособия, читающие курс цифровой обработки изображений в НГТУ и НГУ, исходили из своего понимания важности тех или иных разделов, а также опирались на многолетний научно-исследовательский и педагогический опыт.

При подготовке пособия авторы использовали известные издания в области цифровых методов обработки изображений, а также имеющиеся у них данные об аналогичных дисциплинах, присутствующих в учебных планах ряда высших учебных заведений страны. Среди них — школы в области цифровой обработки изображений Самарского государственного аэрокосмического университета под руководством В.А. Сойфера и В.В. Сергеева и Ульяновского технического университета под руководством К.К. Васильева. Авторы выражают им признательность за то сотрудничество, к числу результатов которого относится и настоящее учебное пособие.





## ГЛАВА 1

### ДИСКРЕТИЗАЦИЯ И КВАНТОВАНИЕ НЕПРЕРЫВНЫХ ИЗОБРАЖЕНИЙ

Очень редко изображения, получаемые в информационных системах, имеют цифровую форму. Поэтому их преобразование к этому виду является обязательной операцией, если предполагается использовать цифровую обработку, передачу, хранение. Как и при одномерных сигналах, данное преобразование включает в себя две процедуры. Первая состоит в замене непрерывного кадра дискретным и обычно называется *дискретизацией*, а вторая выполняет замену непрерывного множества значений яркости множеством квантованных значений и носит название *квантования*. При цифровом представлении каждому из квантованных значений яркости ставится в соответствие двоичное число, чем и достигается возможность ввода изображения в ЭВМ.

Двумерный характер изображения по сравнению с обычными сигналами содержит дополнительные возможности оптимизации цифрового представления с целью сокращения объема получаемых цифровых данных. В связи с этим изучался вопрос о наилучшем размещении уровней квантования, а также об использовании различных растров [1.1...1.3]. Следует, однако, сказать, что в подавляющем большинстве случаев на практике применяют дискретизацию, основанную на использовании прямоугольного растра, и равномерное квантование яркости. Это связано с простотой выполнения соответствующих операций и относительно небольшими преимуществами использования оптимальных преобразований. При использовании прямоугольного растра в окончательном виде цифровое изображение обычно представляет собой матрицу, строки и столбцы которой соответствуют строкам и столбцам изображения.

### 1.1. ДИСКРЕТИЗАЦИЯ НЕПРЕРЫВНЫХ ИЗОБРАЖЕНИЙ

Замену непрерывного изображения дискретным можно выполнить различными способами. Можно, например, выбрать какую-либо систему ортогональных функций и, вычислив коэффициенты представления изображения по этой системе (по этому базису), заменить ими изображение. Многообразие базисов дает возможность образования различных дискретных представлений непрерывного изображения. Однако наиболее употребительной является периодическая дискретизация, в частности, как упоминалось выше, дискретизация с прямоугольным растром. Такой способ дискретизации может рассматриваться как один из вариантов применения ортогонального базиса, использующего в качестве своих элементов сдвинутые  $\delta$ -функции. Далее, следуя [1.1], подробно рассмотрим основные особенности прямоугольной дискретизации.

Пусть  $x_{\text{н}}(t_1, t_2)$  — непрерывное изображение, а  $x(i_1, i_2)$  — соответствующее ему дискретное, полученное из непрерывного путем прямоугольной дискретизации. Это означает, что связь между ними определяется выражением

$$x(i_1, i_2) = x_{\text{н}}(i_1 \Delta t_1, i_2 \Delta t_2), \quad (1.1)$$

где  $\Delta t_1, \Delta t_2$  — соответственно вертикальный и горизонтальный шаги, или интервалы дискретизации. Рис. 1.1 иллюстрирует расположение отсчетов на плоскости  $(t_1, t_2)$  при прямоугольной дискретизации.

Основной вопрос, который возникает при замене непрерывного изображения дискретным, состоит в определении условий, при которых такая замена является полноценной, т.е. не сопровождается потерей информации, содержащейся в непрерывном

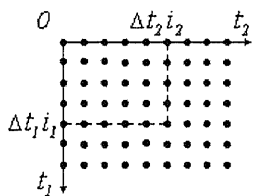


Рис. 1.1. Расположение отсчетов при прямоугольной дискретизации

сигнале. Потери отсутствуют, если, располагая дискретным сигналом, можно восстановить непрерывный. С математической точки зрения, вопрос, таким образом, заключается в восстановлении непрерывного сигнала в двумерных промежутках между узлами, в которых его значения известны, или, иными словами, в осуществлении двумерной интер-

поляции. Ответить на этот вопрос можно, анализируя спектральные свойства непрерывного и дискретного изображений.

Двумерный непрерывный частотный спектр  $X_{\text{н}}(\Omega_1, \Omega_2)$  непрерывного сигнала  $x_{\text{н}}(t_1, t_2)$  определяется двумерным прямым преобразованием Фурье:

$$X_{\text{н}}(\Omega_1, \Omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_{\text{н}}(t_1, t_2) \exp(-j\Omega_1 t_1 - j\Omega_2 t_2) dt_1 dt_2, \quad (1.2)$$

которому отвечает двумерное обратное непрерывное преобразование Фурье:

$$x_{\text{н}}(t_1, t_2) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_{\text{н}}(\Omega_1, \Omega_2) \exp(j\Omega_1 t_1 + j\Omega_2 t_2) d\Omega_1 d\Omega_2. \quad (1.3)$$

Последнее соотношение верно при любых значениях  $t_1, t_2$ , в том числе и в узлах прямоугольной решетки  $t_1 = i_1 \Delta t_1, t_2 = i_2 \Delta t_2$ . Поэтому для значений сигнала в узлах, учитывая (1.1), соотношение (1.3) можно записать в виде:

$$x(i_1, i_2) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_{\text{н}}(\Omega_1, \Omega_2) \exp(j\Omega_1 i_1 \Delta t_1 + j\Omega_2 i_2 \Delta t_2) d\Omega_1 d\Omega_2. \quad (1.4)$$

Обозначим для краткости через  $S(k_1, k_2)$  прямоугольный участок в двумерной частотной области

$$\frac{-\pi + 2\pi k_1}{\Delta t_1} \leq \Omega_1 \leq \frac{\pi + 2\pi k_1}{\Delta t_1}, \quad \frac{-\pi + 2\pi k_2}{\Delta t_2} \leq \Omega_2 \leq \frac{\pi + 2\pi k_2}{\Delta t_2}.$$

Вычисление интеграла в (1.4) по всей частотной области можно заменить интегрированием по отдельным участкам  $S(k_1, k_2)$  и суммированием результатов:

$$x(i_1, i_2) = \frac{1}{4\pi^2} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \iint_{S(k_1, k_2)} X_{\text{н}}(\Omega_1, \Omega_2) \times \\ \times \exp(j\Omega_1 \Delta t_1 i_1 + j\Omega_2 \Delta t_2 i_2) d\Omega_1 d\Omega_2.$$

Выполняя замену переменных по правилу  $\Omega_1 \rightarrow \Omega_1 - 2\pi k_1 / \Delta t_1$ ,  $\Omega_2 \rightarrow \Omega_2 - 2\pi k_2 / \Delta t_2$ , добиваемся независимости области интегрирования от номеров  $k_1$  и  $k_2$ :

$$x(i_1, i_2) = \frac{1}{4\pi^2} \int_{-\frac{\pi}{\Delta t_1}}^{\frac{\pi}{\Delta t_1}} \int_{-\frac{\pi}{\Delta t_2}}^{\frac{\pi}{\Delta t_2}} \sum_{k_1} \sum_{k_2} X_{ii} \left( \Omega_1 + \frac{2\pi k_1}{\Delta t_1}, \Omega_2 + \frac{2\pi k_2}{\Delta t_2} \right) \times \\ \times \exp(j\Omega_1 \Delta t_1 i_1 + j\Omega_2 \Delta t_2 i_2) d\Omega_1 d\Omega_2.$$

Здесь учтено, что  $\exp(-j2\pi ki) = 1$  при любых целых значениях  $k$  и  $i$ . Данное выражение по своей форме очень близко к обратному преобразованию Фурье. Отличие состоит лишь в неправильном виде экспоненциального множителя. Для придания ему необходимого вида введем нормированные частоты  $\omega_1 = \Omega_1 \Delta t_1$ ,  $\omega_2 = \Omega_2 \Delta t_2$  и выполним в соответствии с этим замену переменных. В результате получим:

$$x(i_1, i_2) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{\Delta t_1 \Delta t_2} \sum_{k_1} \sum_{k_2} X_{ii} \left( \frac{\omega_1 + 2\pi k_1}{\Delta t_1}, \frac{\omega_2 + 2\pi k_2}{\Delta t_2} \right) \times \\ \times \exp(j\omega_1 i_1 + j\omega_2 i_2) d\omega_1 d\omega_2. \quad (1.5)$$

Теперь выражение (1.5) имеет форму обратного преобразования Фурье, следовательно, стоящая под знаком интеграла функция

$$X(\omega_1, \omega_2) = \frac{1}{\Delta t_1 \Delta t_2} \sum_{k_1} \sum_{k_2} X_{ii} \left( \frac{\omega_1 + 2\pi k_1}{\Delta t_1}, \frac{\omega_2 + 2\pi k_2}{\Delta t_2} \right) \quad (1.6)$$

является двумерным спектром дискретного изображения. В плоскости ненормированных частот выражение (1.6) имеет вид:

$$X(\Omega_1 \Delta t_1, \Omega_2 \Delta t_2) = \frac{1}{\Delta t_1 \Delta t_2} \sum_{k_1} \sum_{k_2} X_{ii} \left( \Omega_1 + \frac{2\pi k_1}{\Delta t_1}, \Omega_2 + \frac{2\pi k_2}{\Delta t_2} \right). \quad (1.7)$$

Из (1.7) следует, что двумерный спектр дискретного изображения является прямоугольно периодическим с периодами  $2\pi/\Delta t_1$  и  $2\pi/\Delta t_2$  по осям частот  $\Omega_1$  и  $\Omega_2$  соответственно. Спектр дискретного изображения  $X(\Omega_1, \Omega_2)$  образуется в результате суммирования бесконечного количества спектров  $X_{ii}(\Omega_1, \Omega_2)$  непрерывного изображения, отличающихся друг от друга частотными сдвигами  $2\pi/\Delta t_1$  и  $2\pi/\Delta t_2$ . Рис. 1.2 качественно показывает соотношение между двумерными спектрами непрерывного (рис. 1.2, а) и дискретного (рис. 1.2, б) изображений. Сам результат суммирования существенно зависит от значений этих частотных сдвигов, или, иными словами, от выбора интервалов

дискретизации  $\Delta t_1, \Delta t_2$ . Допустим, что спектр непрерывного изображения  $X_{\text{н}}(\Omega_1, \Omega_2)$  отличен от нуля в некоторой двумерной области в окрестности нулевой частоты, т. е. описывается двумерной финитной функцией. Если при этом интервалы дискретизации выбраны так, что  $X_{\text{н}}(\Omega_1, \Omega_2) = 0$  при  $|\Omega_1| \geq \pi/\Delta t_1$ ,  $|\Omega_2| \geq \pi/\Delta t_2$ , то наложения отдельных ветвей при формировании суммы (1.7) происходить не будет. Следовательно, в пределах каждого прямоугольного участка  $S(k_1, k_2)$  от нуля будет отличаться лишь одно слагаемое.

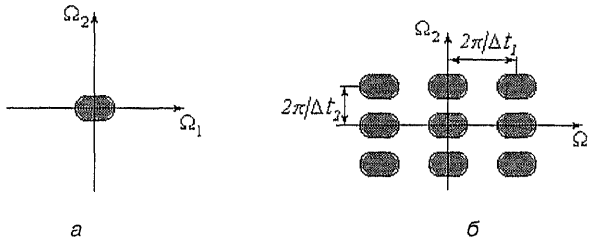


Рис. 1.2. Частотные спектры непрерывного и дискретного изображений

В частности, при  $k_1 = 0$ ,  $k_2 = 0$  имеем:

$$X(\Omega_1, \Omega_2) = \frac{1}{\Delta t_1 \Delta t_2} X_{\text{н}}(\Omega_1, \Omega_2) \quad (1.8)$$

при  $|\Omega_1| \leq \pi/\Delta t_1$ ,  $|\Omega_2| \leq \pi/\Delta t_2$ .

Таким образом, в пределах частотной области  $S(0,0)$  спектры непрерывного и дискретного изображений с точностью до постоянного множителя совпадают. При этом спектр дискретного изображения в данной частотной области содержит полную информацию о спектре непрерывного изображения. Подчеркнем, что такое совпадение имеет место лишь при оговоренных условиях, определяемых правильным выбором интервалов дискретизации. Отметим, что выполнение этих условий согласно (1.8) достигается при достаточно малых значениях интервалов дискретизации  $\Delta t_1, \Delta t_2$ , которые должны удовлетворять требованиям:

$$\Delta t_1 \leq \pi/\Omega_{1\text{гр}}, \quad \Delta t_2 \leq \pi/\Omega_{2\text{гр}}, \quad (1.9)$$

где  $\Omega_{1\text{гр}}$ ,  $\Omega_{2\text{гр}}$  — граничные частоты двумерного спектра.

Соотношение (1.8) определяет способ получения непрерывного изображения  $x_{\text{н}}(t_1, t_2)$  из дискретного  $x(i_1, i_2)$ . Для этого достаточно выполнить двумерную фильтрацию дискретного изображения фильтром низких частот с частотной характеристикой

$$K(j\Omega_1, j\Omega_2) = \begin{cases} \Delta t_1 \Delta t_2 & \text{при } |\Omega_1| \leq \pi/\Delta t_1, \quad |\Omega_2| \leq \pi/\Delta t_2, \\ 0 & \text{при других } \Omega_1, \Omega_2. \end{cases} \quad (1.10)$$

Спектр изображения на его выходе содержит ненулевые компоненты лишь в частотной области  $S(0,0)$  и равняется согласно (1.8) спектру непрерывного изображения  $X_{\text{н}}(\Omega_1, \Omega_2)$ . Это означает, что изображение на выходе идеального фильтра нижних частот совпадает с  $x_{\text{н}}(t_1, t_2)$ .

Таким образом, идеальное интерполяционное восстановление непрерывного изображения выполняется при помощи двумерного фильтра с прямоугольной частотной характеристикой (1.10). Нетрудно записать в явном виде алгоритм восстановления непрерывного изображения. Двумерная импульсная характеристика восстанавливающего фильтра, которую легко получить при помощи обратного преобразования Фурье от (1.10), имеет вид:

$$h(t_1, t_2) = \frac{\sin(\pi t_1 / \Delta t_1)}{\pi t_1 / \Delta t_1} \frac{\sin(\pi t_2 / \Delta t_2)}{\pi t_2 / \Delta t_2}.$$

Продукт фильтрации может быть определен при помощи двумерной свертки входного изображения и данной импульсной характеристики. Представив входное изображение  $x_{\text{вх}}(t_1, t_2)$  в виде двумерной последовательности  $\delta$ -функций

$$x_{\text{вх}}(t_1, t_2) = \sum_{i_1} \sum_{i_2} x(i_1, i_2) \delta(t_1 - i_1 \Delta t_1) \delta(t_2 - i_2 \Delta t_2),$$

после выполнения свертки найдем:

$$x(t_1, t_2) = \sum_{i_1} \sum_{i_2} x(i_1, i_2) \frac{\sin[\pi(t_1 - i_1 \Delta t_1) / \Delta t_1]}{\pi(t_1 - i_1 \Delta t_1) / \Delta t_1} \frac{\sin[\pi(t_2 - i_2 \Delta t_2) / \Delta t_2]}{\pi(t_2 - i_2 \Delta t_2) / \Delta t_2}. \quad (1.11)$$

Полученное соотношение указывает способ точного интерполяционного восстановления непрерывного изображения по известной последовательности его двумерных отсчетов. Согласно этому выражению для точного восстановления в роли интерполирующих функций должны использоваться двумерные функции ви-

да  $(\sin x/x)(\sin y/y)$ . Соотношение (1.11) представляет собой двумерный вариант теоремы Котельникова-Найквиста.

Подчеркнем еще раз, что эти результаты справедливы, если двумерный спектр сигнала является финитным, а интервалы дискретизации достаточно малы. Справедливость сделанных выводов нарушается, если хотя бы одно из этих условий не выполняется. Реальные изображения редко имеют спектры с ярко выраженными граничными частотами. Одна из причин, приводящих к неограниченности спектра, – это ограниченность размеров изображения. Из-за этого при суммировании в (1.7) в каждой из зон  $S(k_1, k_2)$  проявляется действие слагаемых из соседних спектральных зон. При этом точное восстановление непрерывного изображения становится принципиально невозможным. В частности, не приводит к точному восстановлению и использование фильтра с прямоугольной частотной характеристикой.

Условием точного восстановления изображения в промежутках между отсчетами служит использование всех отсчетов дискретного изображения, как это предписывается процедурой (1.11). Это не всегда удобно, часто требуется восстанавливать сигнал в локальной области, опираясь на некоторое небольшое количество имеющихся дискретных значений. В таких случаях целесообразно применять приближенное восстановление при помощи различных интерполирующих функций. Подобная задача возникает, например, при решении проблемы привязки двух изображений, когда из-за геометрических расстройек этих изображений имеющиеся отсчеты одного из них могут соответствовать некоторым точкам, находящимся в промежутках между узлами другого. Решение этой задачи более подробно обсуждается в последующих разделах данного пособия.

Рис. 1.3 иллюстрирует влияние интервалов дискретизации на восстановление изображений. Исходное изображение, представляющее собой отпечаток пальца, приведено на рис. 1.3, *а*, а одно из сечений его нормированного спектра – на рис. 1.3, *б*. Данное изображение является дискретным, а в качестве граничной частоты использовано значение  $\omega_{\text{гп}} = 2\pi \cdot 128$ . Как следует из рис. 1.3, *б*, значение спектра на этой частоте пренебрежимо мало, что гарантирует качественное восстановление. По сути дела, наблюдаемая на рис. 1.3, *а* картина и есть результат восстановления непрерывного изображения, а роль восстанавливающего фильтра выполняет устройство визуализации – монитор или принтер. В этом смысле изображение рис. 1.3, *а* может рассматриваться как непрерывное.

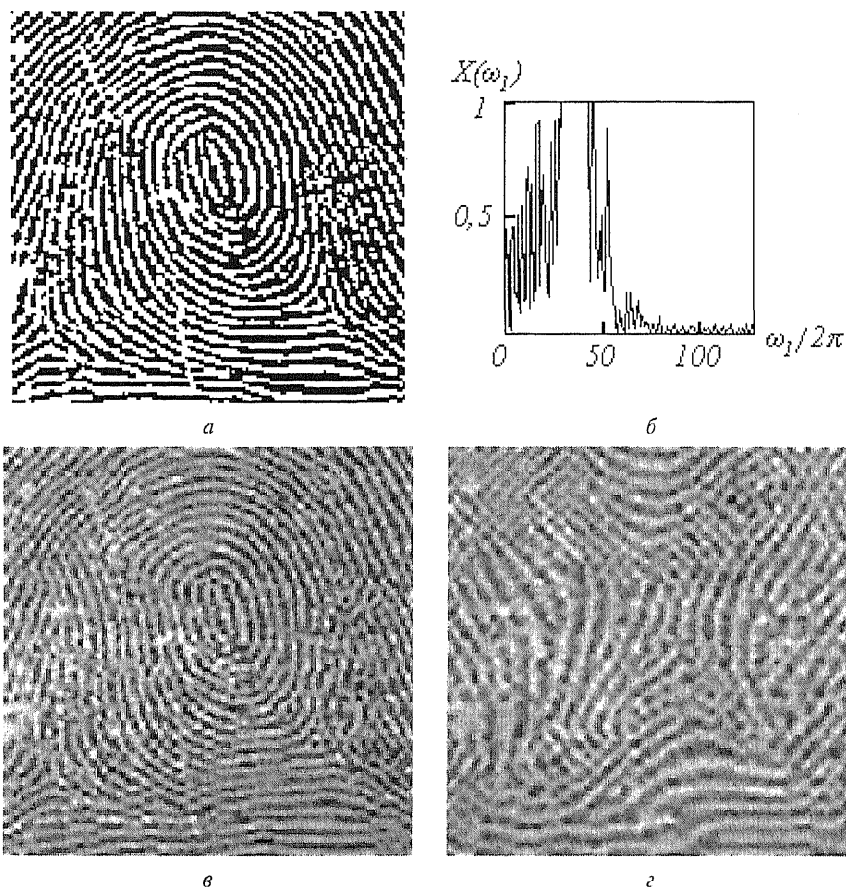


Рис. 1.3. Влияние интервала дискретизации на восстановление изображения «Отпечаток пальца»

Рис. 1.3, в, г показывают последствия от неправильного выбора интервалов дискретизации. При их получении осуществлялась «дискретизация непрерывного» изображения рис. 1.3, а путем прореживания его отсчетов. Рис. 1.3, в соответствует увеличению шага дискретизации по каждой координате в три, а рис. 1.3, г — в четыре раза. Это было бы допустимо, если бы значения граничных частот были ниже в такое же число раз. В действительности, как видно из рис. 1.3, б, происходит нарушение требований (1.9), особенно грубое при четырехкратном прореживании отсчетов. Поэтому восстановленные при помощи алгоритма (1.11) изображения не только оказываются расфокусированными, но и сильно искажают текстуру отпечатка.



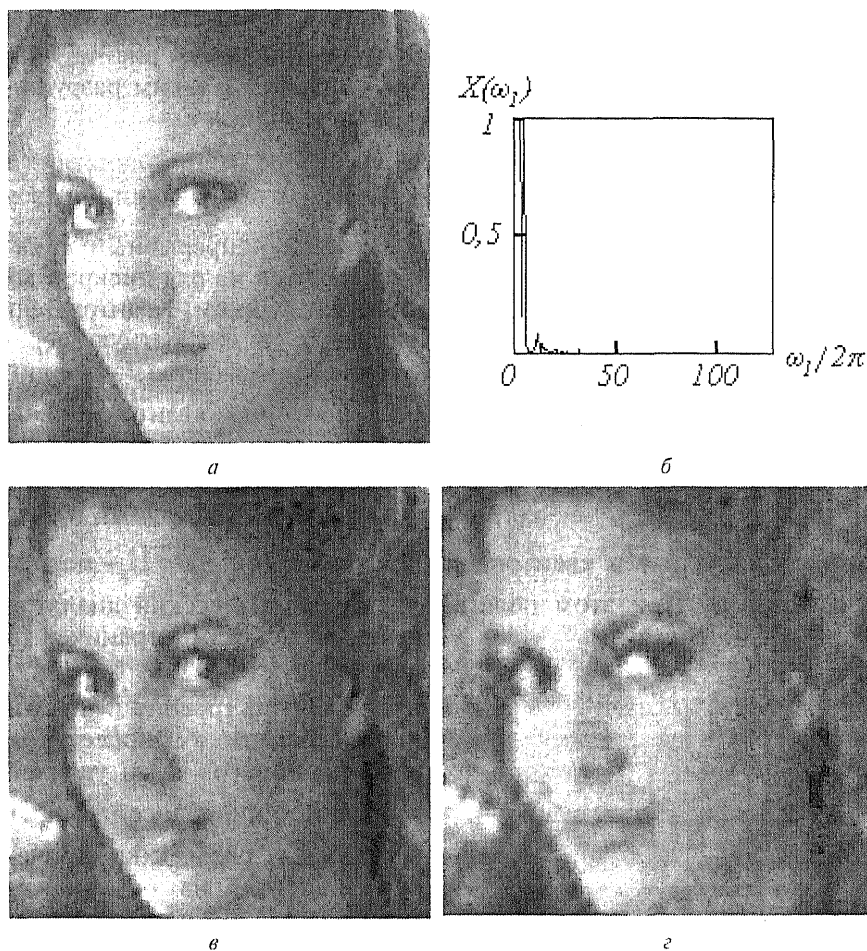


Рис. 1.4. Влияние интервала дискретизации на восстановление изображения «Портрет»

На рис. 1.4 приведена аналогичная серия результатов, полученных для изображения типа «портрет». Последствия более сильного прореживания (в четыре раза на рис. 1.4, в и в шесть раз на рис. 1.4, г) проявляются в основном в потере четкости. Субъективно потери качества представляются менее значительными, чем на рис. 1.3. Это находит свое объяснение в значительно меньшей ширине спектра, чем у изображения отпечатка пальца. Дискретизация исходного изображения соответствует граничной частоте  $\omega_{\text{гп}} = 2\pi \cdot 86$ . Как видно из рис. 1.4, б, это

значение намного превышает истинное значение  $\omega_{igr}$ . Поэтому увеличение интервала дискретизации, иллюстрируемое рис. 1.3, *в, г*, хотя и ухудшает картину, все же не приводит к таким разрушительным последствиям, как в предыдущем примере.

## 1.2. КВАНТОВАНИЕ ИЗОБРАЖЕНИЙ

При цифровой обработке изображений непрерывный динамический диапазон значений яркости делится на ряд дискретных уровней. Эта процедура называется квантованием. Квантователь преобразует непрерывную переменную  $x$  в дискретную переменную  $x_{kv}$ , принимающую конечное множество значений  $\{r_1, \dots, r_L\}$ . Эти значения называются уровнями квантования. В общем случае преобразование выражается ступенчатой функцией (рис. 1.5). Если яркость  $x$  отсчета изображения принадлежит интервалу  $(d_j, d_{j+1}]$  (т.е., когда  $d_j < x \leq d_{j+1}$ ), то исходный отсчет заменяется уровнем квантования  $r_j$ , где  $d_j, j = \overline{1, L+1}$  — пороги квантования. При этом полагается, что динамический диапазон значений яркости ограничен и равен  $[d_1, d_{L+1}]$ .

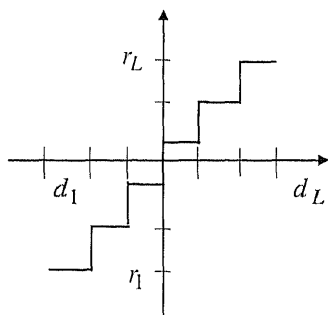


Рис. 1.5. Функция, описывающая квантование

Задача построения квантователя состоит в определении значений порогов  $d_j$  и уровней  $r_j$ . Простейший способ решить эту задачу — разбить динамический диапазон на одинаковые интервалы. Однако такое решение не является наилучшим. Если значения яркости большинства отсчетов изображения сгруппированы, например, в

«темной» области и число уровней  $L$  ограничено, то целесообразно квантовать неравномерно. В «темной» области следует квантовать чаще, а в «светлой» реже. Это позволит уменьшить ошибку квантования  $\varepsilon = x - x_{kv}$ .

Таким образом, задачу построения квантователя можно сформулировать как задачу нахождения оптимальных значений  $d_j$  и  $r_j$ , удовлетворяющих некоторому критерию оптимизации. Обычно при фиксированном числе уровней квантователь опти-

мизируется по критерию минимальной среднеквадратической ошибки

$$E\{\varepsilon^2\} = E\{(x - x_{\text{кв}})^2\} = \min_{\{d_j, r_j\}} \quad (1.12)$$

в предположении, что яркость  $x$  — случайная величина с известной плотностью вероятности  $w(x)$ .

Среднеквадратическая ошибка квантования (1.12)

$$E\{\varepsilon^2\} = \int_{d_1}^{d_{L+1}} (x - x_{\text{кв}})^2 w(x) dx = \sum_{j=1}^L \int_{d_j}^{d_{j+1}} (x - r_j)^2 w(x) dx \quad (1.13)$$

Дифференцируя (1.13) по переменным  $d_j$ ,  $r_j$  и приравнявая производные нулю, получаем  $2L-1$  нелинейных уравнений

$$(d_j - r_j)^2 w(d_j) - (d_j - r_{j-1})^2 w(d_j) = 0, \quad j = \overline{2, L}, \quad (1.14)$$

$$\int_{d_j}^{d_{j+1}} (x - r_j) w(x) dx = 0, \quad j = \overline{1, L}.$$

Надо отметить, что крайние пороги  $d_1$  и  $d_{L+1}$  определяются динамическим диапазоном яркости. Уравнения (1.14) нетрудно привести к виду

$$d_j = (r_j + r_{j-1}) / 2, \quad j = \overline{2, L}, \quad (1.15)$$

$$r_j = \frac{\int_{d_j}^{d_{j+1}} x w(x) dx}{\int_{d_j}^{d_{j+1}} w(x) dx}, \quad j = \overline{1, L}.$$

Из (1.15) следует, что пороги  $d_j$  должны располагаться посередине между двумя соседними уровнями  $r_{j-1}$  и  $r_j$ . Решение этих уравнений можно найти итеративным способом. Оптимальный квантователь, удовлетворяющий критерию (1.12), называется квантователем Ллойда-Макса [1.3, 1.5], а среднеквадратическая ошибка для такого квантователя будет

$$E\{\varepsilon^2\} = E\{x^2\} - \sum_{j=1}^L r_j^2 \int_{d_j}^{d_{j+1}} w(x) dx \quad (1.16)$$

При равномерном распределении яркости нелинейные уравнения (1.15) можно представить в виде [1.3]

$$d_j = d_1 + (j-1)\Delta, \quad j = \overline{2, L},$$

$$r_j = d_j + \Delta/2, \quad j = \overline{1, L},$$

$$\Delta = (d_{L+1} - d_1) / L,$$

а среднеквадратическую ошибку

$$E\{\varepsilon^2\} = \frac{\Delta^2}{12}.$$

В системах цифровой обработки изображений стремятся уменьшить число уровней и порогов квантования, так как от их количества зависит длина двоичного кодового слова, которым представляются проквантованные отсчеты в ЭВМ. Однако при относительно небольшом числе уровней  $L$  на проквантованном изображении появляются ложные контуры. Они особенно заметны на пологих участках изменения яркости вследствие возникновения ее скачков на проквантованном изображении (рис. 1.6).

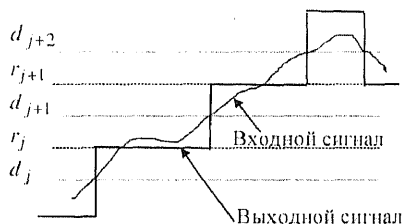
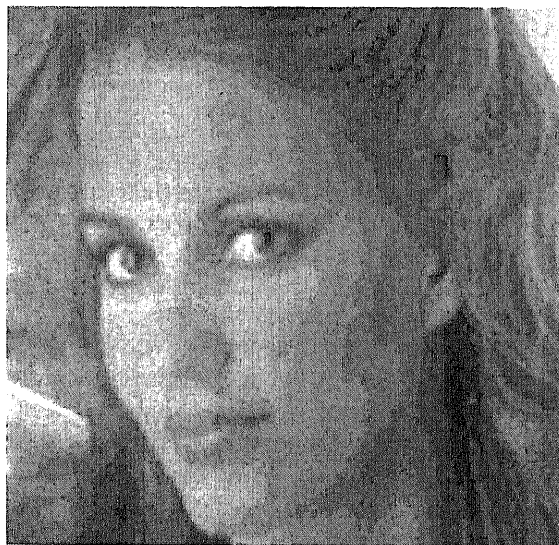


Рис. 1.6. К механизму возникновения ложных контуров

Ложные контуры значительно ухудшают визуальное качество изображения, поскольку зрение человека особенно чувствительно именно к контурам. При равномерном квантовании типичных изображений требуется не менее 64 уровней. На рис. 1.7, а, б приведены результаты равномерного квантования изображения «Портрет» соответственно на 256 и 14 уровней квантования.

В темных частях изображения на рис. 1.7, б заметны ложные контуры. Использование квантователя Ллойда-Макса позволяет существенно снизить их уровень (см. рис. 1.8, где число уровней квантования также равно 14). На рис. 1.9 показана гистограмма яркости изображения «Портрет» при 256 уровнях квантования

*а**б*

*Рис. 1.7.* Результаты равномерного квантования



Рис. 1.8. Результат неравномерного квантования

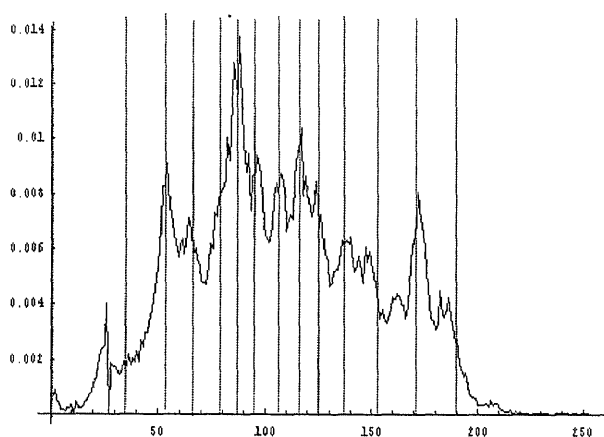


Рис. 1.9. Гистограмма изображения «Портрет»

и отмечены пороги  $d_j$  при  $L + 1 = 15$ . Из рисунка следует, что чаще квантуются те области динамического диапазона, в которых сгруппированы значения яркости отсчетов.

Чтобы избежать неравномерного квантования, которое не может быть выполнено с помощью стандартного АЦП, используют нелинейные преобразования (рис. 1.10). Отсчет  $x$  исходного изображения подвергают нелинейному преобразованию, чтобы плотность распределения вероятностей преобразованных отсчетов  $y$  была равномерной, т.е. выполняют процедуру эквализации, которая подробно описана в главе 2. Затем отсчеты  $y$  квантуют с равномерным шагом и подвергают обратному нелинейному преобразованию.

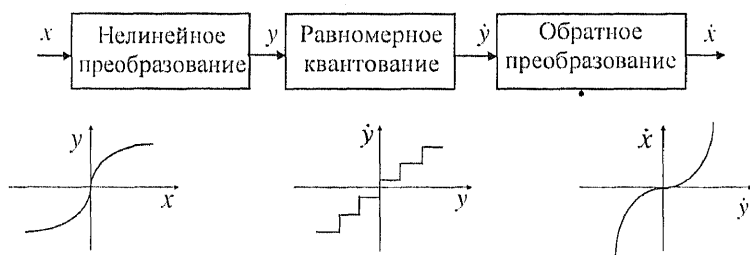


Рис. 1.10. Квантование с предварительным нелинейным преобразованием

Для разрушения ложных контуров Робертс предложил перед равномерным квантованием к отсчетам яркости добавлять шум с равномерной плотностью распределения вероятностей. Добавленный шум переводит одни отсчеты изображения на уровень выше, а другие на уровень ниже. Тем самым разрушаются ложные контуры. Дисперсия добавляемого шума должна быть небольшой, чтобы не привести к искажениям, воспринимаемым как «снег» на изображении, и в то же время достаточной для разрушения ложных контуров. Обычно используют равномерно распределенный шум на интервале  $[-\Delta/2, \Delta/2]$ . Результаты равномерного квантования на 14 и 8 уровней изображения «Портрет» с предварительным добавлением шума приведены на рис. 1.11, а и б. При 8 уровнях квантования добавляемый шум становится слишком заметным, однако ложные контуры разрушены практически полностью.

Еще один метод квантования используется в полиграфии. Это метод формирования растровых бинарных (2-уровневых) изображений

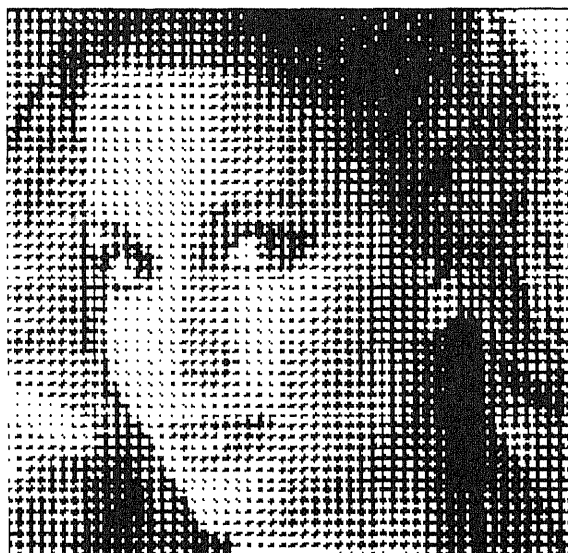
*a**б*

*Рис. 1.11.* Результаты равномерного квантования с предварительным добавлением шума





а



б

Рис. 1.12. Растривание изображений (см. также с. 28)



в

Рис. 1.12. Окончание

из полутоновых. При печати (например, газет или журналов) изображение формируется из белых и черных точек. Для этого все исходное изображение разбивается по пространственным координатам на одинаковые квадратные блоки. Обычно блок содержит  $5 \times 5$  элементов. К каждому отсчету блока добавляется число с соответствующими координатами из матрицы возмущающего сигнала, размеры которой равны размерам блока. Например, в качестве матрицы возмущающего сигнала используют числа [1.5]:

$$\begin{bmatrix} 40 & 60 & 150 & 90 & 10 \\ 80 & 170 & 240 & 200 & 110 \\ 140 & 210 & 250 & 220 & 130 \\ 120 & 190 & 230 & 180 & 70 \\ 20 & 100 & 160 & 50 & 30 \end{bmatrix}.$$

Эта операция повторяется для всех блоков. Получаемое при этом изображение квантуется на два уровня. На рис. 1.12, а показано полутоновое изображение «Портрет» с добавленным возмущаю-

щим сигналом. На рис. 1.12, б, в приведены результаты бинарного квантования изображения «Портрет» с добавленным возмущающим сигналом (рис. 1.12, б) и без него (рис. 1.12, в).

Бинарное растровое изображение обеспечивает значительно лучшее зрительное впечатление, чем обычное бинарное изображение. Передача шкалы яркости при растривании достигается благодаря изменению геометрических размеров белого пятна, наблюдаемого на черном фоне. Если в блоке сгруппировались «светлые» отсчеты, то геометрические размеры белого пятна максимальны и равны размеру блока. При снижении яркости его геометрические размеры также уменьшаются. Глаз человека выполняет локальное усреднение, создавая иллюзию наблюдения полутонового изображения. Процедура растривания особенно эффективна при печати изображений с высоким разрешением, когда одиночное пятно едва различимо глазом.

## ВОПРОСЫ К ГЛАВЕ 1

1.1. Поясните, как проблема дискретизации изображения связана с двумерной интерполяцией.

1.2. Какова структура двумерного частотного спектра дискретного изображения?

1.3. При каких условиях, используя дискретное изображение, можно без потерь восстановить непрерывное?

1.4. Докажите, что двумерный фильтр с прямоугольной частотной характеристикой идеально восстанавливает непрерывное изображение из дискретного.

1.5. Учитывая, что спектры реальных изображений не являются финитными функциями, предложите метод расчета ошибок восстановления, вызванных нарушением требований (1.9).

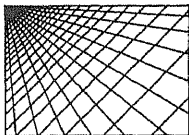
1.6. Поясните, в чем состоит задача поиска оптимального квантователя изображения.

1.7. Чем объясняется стремление использовать небольшое число уровней квантования и что препятствует этой тенденции?

1.8. Поясните механизм возникновения ложных контуров при малом числе уровней квантования изображения.

1.9. Каким образом добавление шума к изображению позволяет ослабить эффект ложных контуров, возникающих при квантовании? Как при этой процедуре следует выбирать характеристики шума?

1.10. С какой целью применяют нелинейное предвысказание изображения при его равномерном квантовании? Какой должна быть характеристика нелинейного элемента?



## ГЛАВА 2

### **УЛУЧШЕНИЕ ВИЗУАЛЬНОГО КАЧЕСТВА ИЗОБРАЖЕНИЙ ПУТЕМ ПОЭЛЕМЕНТНОГО ПРЕОБРАЗОВАНИЯ**

В большом числе информационных систем применяется представление результатов обработки данных в виде изображения, выводимого на экран для использования наблюдателем. Процедуру, обеспечивающую такое представление, называют визуализацией. Желательно при помощи обработки придать выводимому изображению такие качества, благодаря которым его восприятие человеком было бы по возможности комфортным. Часто бывает полезным подчеркнуть, усилить какие-то черты, особенности, нюансы наблюдаемой картины с целью улучшения ее субъективного восприятия.

Последнее — субъективность восприятия — сильно усложняет применение формализованного подхода в достижении данных целей. Поэтому при обработке изображений для визуализации получили распространение методы, в которых часто отсутствуют строгие математические критерии оптимальности. Их заменяют качественные представления о целесообразности той или иной обработки, опирающиеся на субъективные оценки результатов.

Подавляющее большинство процедур обработки для получения результата в каждой точке кадра привлекает входные данные из некоторого множества точек исходного изображения, окружающих обрабатываемую точку. Однако имеется группа процедур, где осуществляется так называемая поэлементная обработка. Здесь результат обработки в любой точке кадра зависит только от значения входного изображения в этой же точке. Очевидное достоинство таких процедур — их предельная простота. Вместе с тем многие из них приводят к субъективному улучшению визуального качества. Этим определяется внимание, которое уделяют

поэлементным процедурам. Не преувеличивая их роли, отметим, что очень часто поэлементная обработка применяется как заключительный этап при решении более сложной задачи обработки изображения.

Сущность поэлементной обработки изображений сводится к следующему. Пусть  $x(i, j) = x_{i,j}$ ,  $y(i, j) = y_{i,j}$  — значения яркости исходного и получаемого после обработки изображений соответственно в точке кадра, имеющей декартовы координаты  $i$  (номер строки) и  $j$  (номер столбца). Поэлементная обработка означает, что существует функциональная однозначная зависимость между этими яркостями

$$y_{i,j} = f_{i,j}(x_{i,j}), \quad (2.1)$$

позволяющая по значению исходного сигнала определить значение выходного продукта. В общем случае, как это учтено в данном выражении, вид или параметры функции  $f_{i,j}(\cdot)$ , описывающей обработку, зависят от текущих координат. При этом обработка является *неоднородной*. Однако в большинстве практически применяемых процедур используется *однородная* поэлементная обработка. В этом случае индексы  $i$  и  $j$  в выражении (2.1) могут отсутствовать, а зависимость между яркостями исходного и обработанного изображений описывается функцией

$$y = f(x), \quad (2.2)$$

одинаковой для всех точек кадра.

Цель данной главы состоит в изучении наиболее распространенных процедур поэлементной обработки.

## 2.1. ЛИНЕЙНОЕ КОНТРАСТИРОВАНИЕ ИЗОБРАЖЕНИЯ

Задача контрастирования связана с улучшением согласования динамического диапазона изображения и экрана, на котором выполняется визуализация. Если для цифрового представления каждого отсчета изображения отводится 1 байт (8 бит) запоминающего устройства, то входной или выходной сигналы могут принимать одно из 256 значений. Обычно в качестве рабочего используется диапазон 0...255; при этом значение 0 соответствует при визуализации уровню черного, а значение 255 — уровню белого. Предположим, что минимальная и максимальная яркости исходного изображения равны  $x_{\min}$  и  $x_{\max}$  соответственно. Если

эти параметры или один из них существенно отличаются от граничных значений яркостного диапазона, то визуализированная картина выглядит как ненасыщенная, неудобная, утомляющая при наблюдении. Пример такого неудачного представления приведен на рис. 2.1, *а*, где диапазон яркостей имеет границы  $x_{\min} = 180$ ,  $x_{\max} = 240$ .

При линейном контрастировании используется линейное преобразование вида

$$y = ax + b, \quad (2.3)$$

параметры которого  $a$  и  $b$  определяются желаемыми значениями минимальной  $y_{\min}$  и максимальной  $y_{\max}$  выходной яркости.

Решив систему уравнений

$$\begin{cases} y_{\min} = ax_{\min} + b, \\ y_{\max} = ax_{\max} + b \end{cases}$$

относительно параметров преобразования  $a$  и  $b$ , нетрудно привести (2.3) к виду

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}(y_{\max} - y_{\min}) + y_{\min}.$$

Результат линейного контрастирования исходного изображения, представленного на рис. 2.1, *а*, приведен на рис. 2.1, *б* при  $y_{\min} = 0$  и  $y_{\max} = 255$ . Сравнение двух изображений свидетельствует о значительно лучшем визуальном качестве обработанного изображения. Улучшение связано с использованием после контрастирования полного динамического диапазона экрана, что отсутствует у исходного изображения.

## 2.2. СОЛЯРИЗАЦИЯ ИЗОБРАЖЕНИЯ

При данном виде обработки преобразование (2.2) имеет вид [2.1]:

$$y = kx (x_{\max} - x),$$

где  $x_{\max}$  — максимальное значение исходного сигнала, а  $k$  — константа, позволяющая управлять динамическим диапазоном преобразованного изображения. Функция, описывающая данное преобразование, является квадратичной параболой, ее график при  $k=1$  показан на рис. 2.2. При  $y_{\max} = x_{\max}$  динамические диапазоны изображений совпадают, что может быть достигнуто

*а**б*

Рис. 2.1. Пример линейного контрастирования

при  $k = 4/x_{\max}$ . Другой путь нормализации динамического диапазона может состоять в применении одного из методов контрастирования, например, описанного выше линейного контрастирования.

Как следует из рис. 2.2, смысл соляризации заключается в том, что участки исходного изображения, имеющие уровень белого или близкий к нему уро-

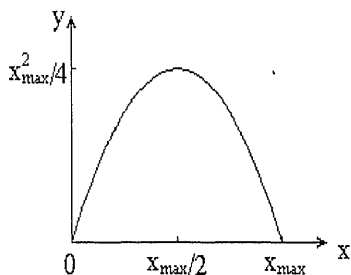


Рис. 2.2. Функция, описывающая соляризацию

вень яркости, после обработки имеют уровень черного. При этом сохраняют уровень черного и участки, имеющие его на исходном изображении. Уровень же белого на выходе приобретают участки, имеющие на входе средний уровень яркости (уровень серого). Пример применения соляризации приведен на рис. 2.3. На рис. 2.3, а показано исходное изображение, а на рис. 2.3, б — результат его соляризации. На втором этапе

обработки здесь применено линейное контрастирование при  $y_{\min} = 0$  и  $y_{\max} = 255$ . Обработка повышает четкость деталей изображения: улучшены изображения глаз, увеличен контраст на переходе «лицо — волосы» и т. д.

### 2.3. ПРЕПАРИРОВАНИЕ ИЗОБРАЖЕНИЯ

Препазирование представляет собой целый класс поэлементных преобразований изображений. Характеристики используемых на практике процедур препазирования приведены на рис. 2.4. Остановимся на описании некоторых из них.

Преобразование с пороговой характеристикой (рис. 2.4, а) превращает полутоновое изображение, содержащее все уровни яркости, в бинарное, точки которого имеют яркости  $y = 0$  или  $y = y_{\max}$ . Такая операция, называемая иногда бинаризацией или бинарным квантованием, может быть полезной, когда для наблюдателя важны очертания объектов, присутствующих на изображении, а детали, содержащиеся внутри объектов или внутри фона, не представляют интереса. Основная проблема при проведении такой обработки — это определение порога  $x_0$ , сравнение



*а**б*

*Рис. 2.3. Пример соляризации*

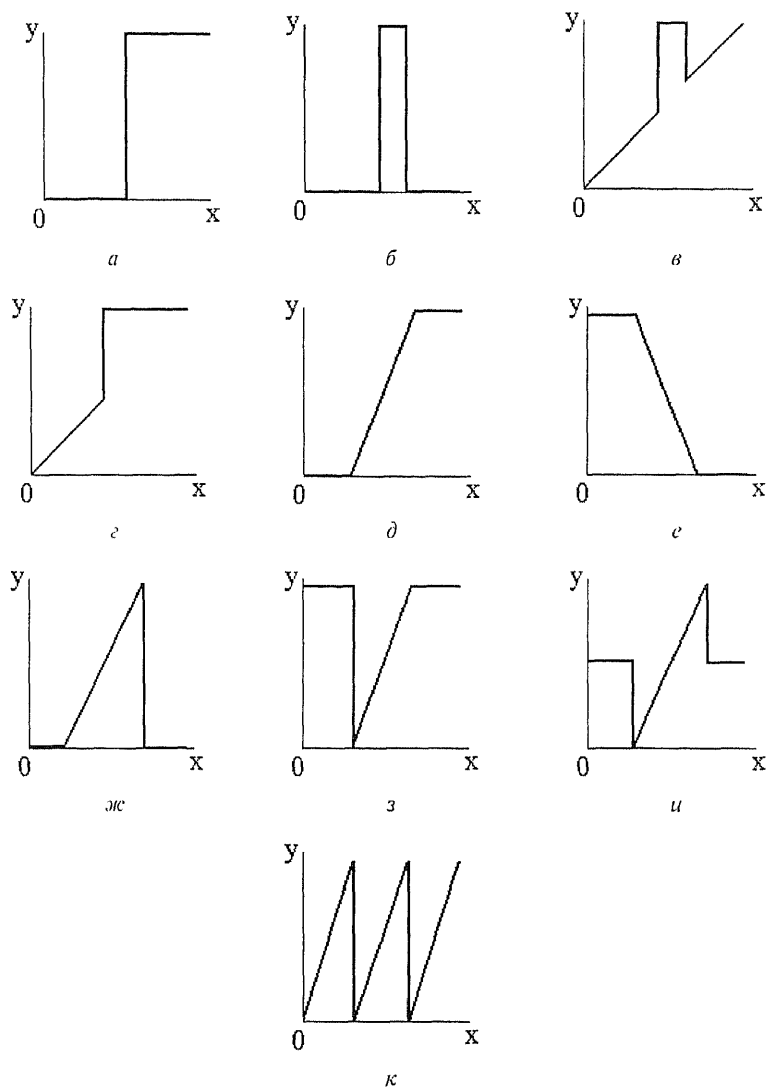


Рис. 2.4. Примеры преобразований, используемых при препарировании

с которым яркости исходного изображения позволяет определить значение выходного изображения в каждой его точке. Наиболее оправданным для математического описания изображения является применение теории вероятностей, случайных процессов и случайных полей. При этом определение оптимального порога бинарного квантования представляет собой статистическую зада-

чу. Статистическому подходу к обработке изображений в последующих разделах уделяется значительное внимание, в том числе и при решении задачи разделения точек изображения на два класса — так называемой бинарной сегментации. Здесь же ограничимся обсуждением частного, но практически важного случая. Иногда при обработке приходится иметь дело с изображениями, хранимыми как полутоновые, но по своему содержанию мало отличающимися от бинарных. К ним относятся текст, штриховые рисунки, чертежи, изображение отпечатка пальца, пример которого приведен на рис. 2.6, а. Плотность вероятности  $w(x)$ , описывающая распределение яркости такого изображения, может содержать два хорошо разделяющихся пика.

Интуитивно понятно, что порог бинарного квантования следует выбирать посередине провала между этими пиками, как это показано на рис. 2.5. Замена исходного полутонового изображения *бинарным препаратом* решает две основные задачи. Во-первых, достигается большая наглядность при визуальном восприятии, чем у исходного изображения. Во-вторых, ощутимо сокращается объем запоминающего устройства для хранения изображения, поскольку бинарный препарат для записи каждой точки бинарного изображения требует лишь 1 бит памяти, в то время как полутоновое изображение для решения той же задачи при наиболее часто применяемом формате представления — 8 бит. Пример бинаризации изображения отпечатка пальца приведен на рис. 2.6, б.

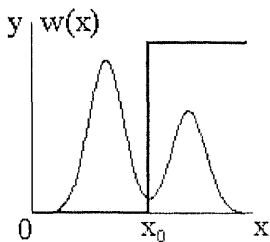


Рис. 2.5. К выбору порога бинарного квантования

Смысл других преобразований, представленных на рис. 2.4, нетрудно понять, рассматривая их характеристики. Например, преобразование рис. 2.4, б выполняет яркостный срез изображения, выделяя те его участки, где яркость соответствует выделенному интервалу. При этом остальные участки оказываются полностью «погашенными» (имеют яркость, соответствующую уровню черного). Перемещая выделенный интервал по яркостной шкале и изменяя его ширину, можно детально исследовать содержание картины.

Преобразование, приведенное на рис. 2.4, ж, также позволяет повысить детальность наблюдаемой картины в выбранном диапазоне яркостей, однако в отличие от предыдущего здесь выход-

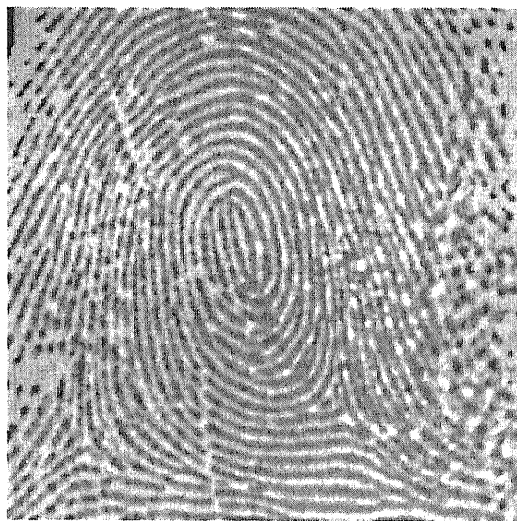
*а**б*

Рис. 2.6. Пример бинаризации изображения

ное изображение использует полный динамический диапазон. По существу, это преобразование представляет собой линейное контрастирование, применяемое к *избранному диапазону* входного изображения. Как и в предыдущем варианте, участки, не попавшие в этот диапазон, образуют после препарирования черный фон.

Иногда наглядность изображения повышается использованием преобразования типа пилообразного контрастирования (рис. 2.4, *к*). При этом различные яркостные диапазоны одновременно подвергаются локальному яркостному контрастированию. Однако необходимо иметь в виду, что данное преобразование, как и некоторые другие, может сопровождаться появлением ложных контуров на получаемом препарате.

Аналогично можно качественно рассмотреть и остальные процедуры препарирования, показанные на рис. 2.4.

На рис. 2.7 приведены результаты эксперимента, в котором к аэроснимку участка Земли (рис. 2.7, *а*) применялись преобразования типа «Пороговая обработка» (рис. 2.7, *б*) и «Пилообразное контрастирование» (рис. 2.7, *в*). Первое приводит к выявлению границ отдельных участков, создавая общее интегральное представление о наблюдаемой сцене. Второе, наоборот, дает возможность наблюдать мелкие детали на всех участках изображения. Сочетание двух таких возможностей может оказаться полезным наблюдателю.

В заключение отметим, что препарирование часто используется и в автоматических системах обработки визуальной информации, поскольку подготавливаемый при этом препарат может содержать всю информацию, необходимую для последующей (вторичной) обработки. Например, если при наблюдении из космоса требуется автоматически обнаружить на изображении некоторый объект, имеющий известную конфигурацию, то для этого может быть достаточно бинарного препарата, передающего эту конфигурацию.

## 2.4. ПРЕОБРАЗОВАНИЕ ГИСТОГРАММ, ЭКВАЛИЗАЦИЯ

При всех поэлементных преобразованиях происходит изменение закона распределения вероятностей, описывающего изображение. Рассмотрим механизм этого изменения на примере произвольного преобразования с монотонной характеристикой, описываемой функцией  $y = f(x)$  (рис. 2.8), имеющей однозначную обратную функцию  $x = \varphi(y)$ . Предположим, что случайная

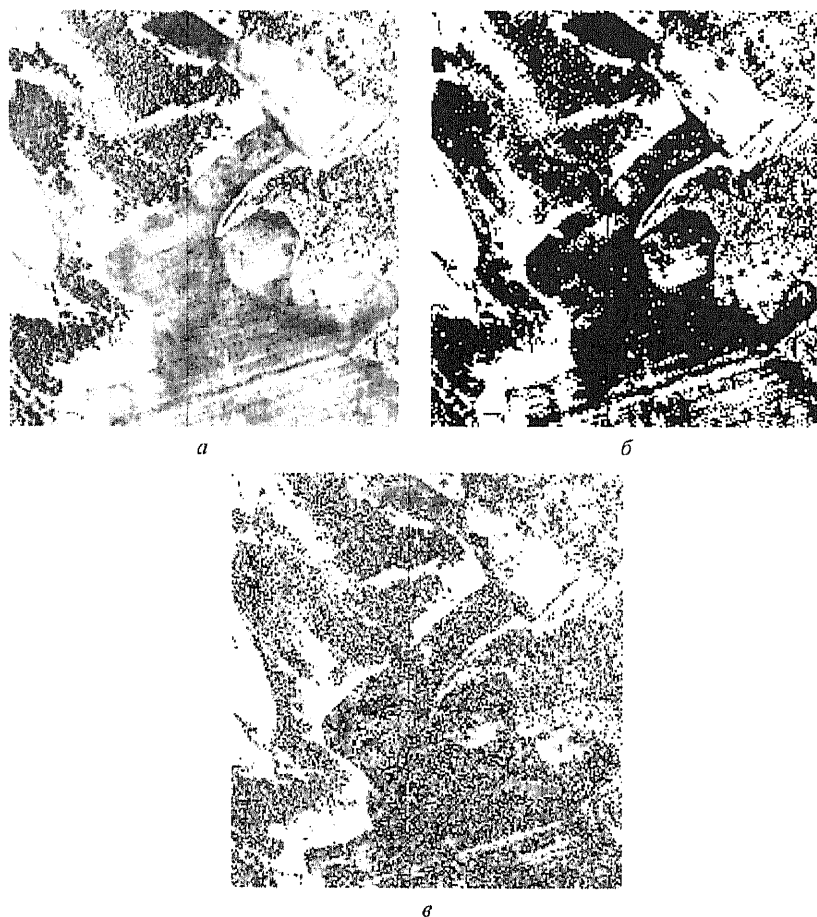


Рис. 2.7. Примеры препарирования изображения

величина  $x$  подчиняется плотности вероятности  $w_x(x)$ . Пусть  $\Delta x$  — произвольный малый интервал значений случайной величины  $x$ , а  $\Delta y$  — *соответствующий* ему интервал преобразованной случайной величины  $y$ . Попадание величины  $x$  в интервал  $\Delta x$  влечет за собой попадание величины  $y$  в интервал  $\Delta y$ , что означает вероятностную эквивалентность этих двух событий. Поэтому, учитывая малость обоих интервалов, можно записать приближенное равенство:

$$w_x(x)|\Delta x| \approx w_y(y)|\Delta y|,$$

где модули учитывают зависимость вероятностей от *абсолютных длин* интервалов (и независимость от *знаков* приращений  $\Delta x$  и  $\Delta y$ ). Вычисляя отсюда плотность вероятности преобразованной величины, подставляя вместо  $x$  его выражение через обратную функцию и выполняя предельный переход при  $\Delta x \rightarrow 0$  (и, следовательно,  $\Delta y \rightarrow 0$ ), получаем:

$$w_y(y) = w_x(\varphi(y)) \left| \frac{d\varphi(y)}{dy} \right|. \quad (2.4)$$

Это выражение позволяет вычислить плотность вероятности продукта преобразования, которая, как видно, *не совпадает* с плотностью распределения исходной случайной величины. Ясно, что существенное влияние на плотность  $w_y(y)$  оказывает выполняемое преобразование, поскольку в (2.4) входят обратная функция и ее производная.

Соотношения становятся несколько сложнее, если преобразование описывается не взаимно однозначной функцией [2.2]. Примером такой более сложной характеристики с неоднозначной обратной функцией может служить пилообразная характеристика рис. 2.4, к. Однако в общем смысл вероятностных преобразований при этом не изменяется.

Все описанные в данной главе поэлементные преобразования изображений можно рассмотреть с точки зрения изменения плотности вероятности, описываемого выражением (2.4). Очевидно, что ни при одном из них плотность вероятности выходного продукта не будет совпадать с плотностью вероятности исходного изображения (за исключением, конечно, тривиального преобразования  $y = f(x) \equiv x$ ). Нетрудно убедиться, что при линейном контрастировании сохраняется *вид* плотности вероятности, однако в общем случае, т.е. при произвольных значениях параметров линейного преобразования, изменяются параметры плотности вероятности преобразованного изображения.

Определение вероятностных характеристик изображений, прошедших нелинейную обработку, является *прямой* задачей анализа. При решении практических задач обработки изображе-

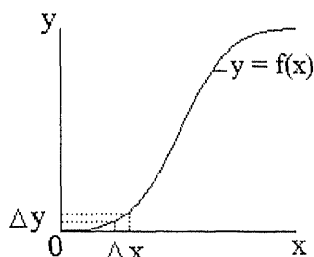


Рис. 2.8. Нелинейное преобразование случайной величины

ний может быть поставлена *обратная задача*: по известному виду плотности вероятности  $w_x(x)$  и *желаемому* виду  $w_y(y)$  определить *требуемое* преобразование  $y = f(x)$ , которому следует подвергнуть исходное изображение. В практике цифровой обработки изображений часто полезный результат получают преобразованием изображения к равновероятному распределению [2.3]. В этом случае

$$w_y(y) = \begin{cases} 1/(y_{\max} - y_{\min}), & \text{при } y_{\min} \leq y \leq y_{\max}, \\ 0, & \text{при других } y, \end{cases} \quad (2.5)$$

где  $y_{\min}$  и  $y_{\max}$  – минимальное и максимальное значения яркости преобразованного изображения. Определим характеристику преобразователя, решающего данную задачу. Пусть  $x$  и  $y$  связаны функцией (2.2), а  $F_x(x)$  и  $F_y(y)$  – интегральные законы распределения входной и выходной величин. Учитывая (2.5), найдем:

$$F_y(y) = \int_{y_{\min}}^y w_y(y) dy = \frac{y - y_{\min}}{y_{\max} - y_{\min}}.$$

Подставляя это выражение в условие вероятностной эквивалентности

$$F_x(x) = F_y(y),$$

после простых преобразований получаем соотношение

$$y = (y_{\max} - y_{\min}) F_x(x) + y_{\min}, \quad (2.6)$$

представляющее собой характеристику (2.2) в решаемой задаче. Согласно (2.6) исходное изображение проходит нелинейное преобразование, характеристика которого  $F_x(x)$  определяется интегральным законом распределения самого исходного изображения. После этого результат приводится к заданному динамическому диапазону при помощи операции линейного контрастирования.

Аналогичным образом могут быть получены решения других подобных задач, в которых требуется привести законы распределения изображения к заданному виду. В [2.4] имеется таблица таких преобразований. Одно из них, так называемая *гиперболизация распределения*, предполагает приведение плотности вероятности преобразованного изображения к гиперболическому виду:



$$w_y(y) = \begin{cases} \frac{1}{y(\ln y_{\max} - \ln y_{\min})}, & \text{при } y_{\min} \leq y \leq y_{\max}, \\ 0, & \text{при других } y. \end{cases} \quad (2.7)$$

Если учесть, что сетчатка глаза логарифмирует яркость входного потока, то итоговая плотность вероятности оказывается равномерной. Таким образом, отличие от предыдущего примера заключается в учете физиологических свойств зрения. Можно показать, что изображение с плотностью вероятности (2.7) получается на выходе нелинейного элемента с характеристикой

$$y = y_{\min} \left( \frac{y_{\max}}{y_{\min}} \right)^{F_x(x)}, \quad (2.8)$$

также определяемой интегральным законом распределения исходного изображения.

Таким образом, преобразование плотности вероятности предполагает знание интегрального распределения для исходного изображения. Как правило, достоверные сведения о нем отсутствуют. Использование для рассматриваемых целей аналитических аппроксимаций также малоприменимо, поскольку их небольшие отклонения от истинных распределений могут приводить к существенному отличию результатов от требуемых. Поэтому в практике обработки изображений преобразование распределений выполняют в два этапа.

На первом этапе измеряют *гистограмму* исходного изображения. Для цифрового изображения, шкала яркостей которого, например, принадлежит целочисленному диапазону 0...255, гистограмма представляет собой таблицу из 256 чисел. Каждое из них показывает количество точек в кадре, имеющих данную яркость. Разделив все числа этой таблицы на общий размер выборки, равный числу используемых точек изображения, получают *оценку* распределения вероятностей яркости изображения. Обозначим эту оценку  $w_x^*(j)$ ,  $0 \leq j \leq 255$ . Тогда оценка интегрального распределения получается по формуле

$$F_x^*(j) = \sum_{i=0}^j w_x^*(i).$$

На втором этапе выполняют само нелинейное преобразование (2.2), обеспечивающее необходимые свойства выходного изобра-

жения. При этом вместо *неизвестного* истинного интегрального распределения используется его оценка, основанная на гистограмме. С учетом этого все методы поэлементного преобразования изображений, целью которых является видоизменение законов распределения, получили название *гистограммных методов*. В частности, преобразование, при котором выходное изображение имеет равномерное распределение, называется *эквализацией (выравниванием) гистограмм*.

Отметим, что процедуры преобразования гистограмм могут применяться как к изображению в целом, так и к отдельным его фрагментам. Последнее может быть полезным при обработке *нестационарных* изображений, содержание которых существенно различается по своим характеристикам на различных участках. В этом случае большего эффекта можно добиться, применяя гистограммную обработку к отдельным участкам.

Использование соотношений (2.4)...(2.8), справедливых для изображений с непрерывным распределением яркости, является не вполне корректным для цифровых изображений. Необходимо иметь в виду, что в результате обработки не удастся получить идеальное распределение вероятностей выходного изображения, поэтому полезно проводить контроль его гистограммы.

На рис. 2.9 приведен пример эквализации, выполненной в соответствии с изложенной методикой. Характерной чертой многих изображений, получаемых в реальных изображающих системах, является значительный удельный вес темных участков и сравнительно малое число участков с высокой яркостью. Эквализация призвана откорректировать картину, выравнивая интегральные площади участков с различными яркостями. Сравнение исходного (рис. 2.9, *а*) и обработанного (рис. 2.9, *б*) изображений показывает, что происходящее при обработке перераспределение яркостей улучшает визуальное восприятие.

## 2.5. ПРИМЕНЕНИЕ ТАБЛИЧНОГО МЕТОДА ПРИ ПОЭЛЕМЕНТНЫХ ПРЕОБРАЗОВАНИЯХ ИЗОБРАЖЕНИЙ

При поэлементных преобразованиях изображений вычисления по формуле (2.2) должны быть выполнены для всех точек исходного изображения. Даже при не очень трудоемких вычислениях в соответствии с видом применяемого преобразования общий объем вычислительной работы, выполняемой ЭВМ, может оказаться значительным. В тех же задачах, где функция  $f(x)$  в (2.2) предполагает трудоемкие вычисления (умножения,

*а**б*

Рис. 2.9. Пример эквализации изображения:  
*а* — исходное изображение; *б* — результат обработки

деления, расчеты тригонометрических выражений, степенных и других функций), применение прямого метода преобразования яркостей может оказаться вообще неприемлемым. Неудобным является и то, что время выполнения обработки зависит от ее вычислительной сложности. От этих недостатков избавлен табличный метод, получивший широкое распространение в практике цифровой обработки изображений.

Сущность табличного метода состоит в том, что путем *предварительного расчета* создается таблица функции  $y = f(x)$  (табл. 2.1).

Таблица 2.1

Табличное задание функции  $y = f(x)$ 

|            |       |       |   |   |   |   |   |           |       |
|------------|-------|-------|---|---|---|---|---|-----------|-------|
| $x$        | $x_1$ | $x_2$ | . | . | . | . | . | $x_{n-1}$ | $x_n$ |
| $y = f(x)$ | $y_1$ | $y_2$ | . | . | . | . | . | $y_{n-1}$ | $y_n$ |

При обработке изображения вместо вычислений пользуются готовыми результатами этой таблицы. При этом значение входной яркости  $x$  служит для определения *номера столбца*, из которого должна быть считана величина преобразованного сигнала  $y$ . Выполнение этой несложной операции по сравнению с вычислением на каждом шаге значения  $y$  позволяет сделать обработку достаточно технологичной, причем ее продолжительность оказывается независимой от сложности преобразования. Необходимо иметь в виду, что все реальные таблицы, которые могут быть записаны в оперативной памяти ЭВМ, имеют ограниченную длину. Если количество значений входного сигнала превышает размеры таблицы, то при попадании значения  $x$  между точками, содержащимися в ней, приходится применять интерполяцию — приближенное определение недостающих значений функции  $y = f(x)$  по имеющимся ее соседним значениям. Часто для этой цели используется *линейная интерполяция*, при которой на участке между заданными узлами функция заменяется отрезком прямой.

Вместе с тем, если исходное изображение представлено в целочисленной записи в диапазоне 0..255, то размеры полной таблицы, содержащей все эти значения, вполне приемлемы для ее хранения. Кроме того, в этом случае значение яркости исходного сигнала служит *адресом*, определяющим номер столбца в таблице. Обработка с использованием этого метода оказывается очень удобной и быстродействующей.

**ВОПРОСЫ К ГЛАВЕ 2**

2.1. В чем состоит сущность поэлементной обработки изображений?

2.2. Как определяются параметры преобразования изображения при его линейном контрастировании?

2.3. Поясните механизм действия поэлементных преобразований, применяемых при препарировании изображения (рис. 2.4).

2.4. Каков механизм появления ложных контуров при применении пилообразного контрастирования? При каких еще процедурах могут возникать ложные контуры?

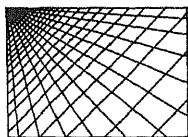
2.5. Докажите, что применение линейного контрастирования не изменяет *вида* плотности вероятности изображения. Как при этом изменяются параметры плотности вероятности?

2.6. Докажите справедливость преобразования (2.7) для получения гиперболического распределения (2.6).

2.7. Докажите, что применение гиперболизации распределения яркости приводит к равновероятному распределению сигнала на выходе сетчатки глаза, если учесть ее логарифмирующее воздействие на входной свет.

2.8. Поясните, почему при эквализации изображения не удастся привести гистограмму к идеальному равномерному виду.

2.9. В чем состоит сущность и каковы достоинства табличного метода поэлементного преобразования изображений?



## ГЛАВА 3

### ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ

Обычно изображения, сформированные различными информационными системами, искажаются действием помех. Это затрудняет как их визуальный анализ человеком-оператором, так и автоматическую обработку в ЭВМ. При решении некоторых задач обработки изображений в роли помех могут выступать те или иные компоненты самого изображения. Например, при анализе космического снимка земной поверхности может стоять задача определения границ между ее отдельными участками: лесом и полем, водой и сушей и т.п. С точки зрения этой задачи, отдельные детали изображения внутри разделяемых областей являются помехой.

Ослабление действия помех достигается фильтрацией. При фильтрации яркость (сигнал) каждой точки исходного изображения, искаженного помехой, заменяется некоторым другим значением яркости, которое признается в наименьшей степени искаженным помехой. Что может послужить основой для таких решений? Изображение часто представляет собой двумерную функцию пространственных координат, которая изменяется по этим координатам медленнее (иногда значительно медленнее), чем помеха, также являющаяся двумерной функцией. Это позволяет при оценке полезного сигнала в каждой точке кадра *принять во внимание* некоторое множество соседних точек, воспользовавшись определенной *похожестью* сигнала в этих точках. В других случаях, наоборот, признаком полезного сигнала являются резкие перепады яркости. Однако, как правило, частота этих перепадов относительно невелика, так что на значительных промежутках между ними сигнал либо постоянен, либо изменяется медленно. И в этом случае свойства сигнала проявляются не только при наблюдении его в локальной точке, но и при анализе

ее окрестности. Заметим, что понятие окрестности является достаточно условным. Окрестность может быть образована лишь ближайшими по кадру соседями, но могут быть окрестности, содержащие сильно удаленные точки кадра. В этом последнем случае, конечно, степень влияния далеких и близких точек на решения, принимаемые фильтром в данной точке кадра, будет совершенно различной.

Таким образом, идеология фильтрации основывается на рациональном использовании данных как из рабочей точки, так и из ее окрестности. В этом проявляется существенное отличие фильтрации от рассмотренных выше поэлементных процедур: фильтрация не может быть поэлементной процедурой обработки изображений.

Задача заключается в том, чтобы найти такую рациональную вычислительную процедуру, которая позволяла бы достигать наилучших результатов. Общепринято при решении этой задачи опираться на использование вероятностных моделей изображения и помехи, а также на применение статистических критериев оптимальности. Причины этого понятны – это и случайный характер как информационного сигнала, так и помехи, это и стремление получить минимальное *в среднем* отличие результата обработки от идеального сигнала. Многообразие методов и алгоритмов связано с большим разнообразием сюжетов, которые приходится описывать различными математическими моделями. Кроме того, применяются различные критерии оптимальности, что также ведет к разнообразию методов фильтрации. Наконец, даже при совпадении моделей и критериев очень часто из-за математических трудностей не удастся найти оптимальную процедуру. Сложность нахождения точных решений порождает различные варианты приближенных методов и процедур.

### 3.1. ОПТИМАЛЬНАЯ ЛИНЕЙНАЯ ФИЛЬТРАЦИЯ. УРАВНЕНИЕ ВИНЕРА – ХОПФА

Пусть  $x_{ij}$  – значение яркости изображения полезного сигнала на пересечении  $i$ -й строки и  $j$ -го столбца, а наблюдаемое изображение описывается моделью

$$y_{i,j} = f(x_{i,j}, n_{i,j}), \quad i = \overline{0, I-1}, \quad j = \overline{0, J-1}. \quad (3.1)$$

Здесь  $n_{ij}$  – значение помехи в точке с координатами  $(i, j)$ ,  $f()$  – функция, описывающая взаимодействие сигнала и помехи, а  $I$  и  $J$  – соответственно число строк и столбцов в кадре.

В дальнейшем будем придерживаться принятой при цифровой обработке изображений декартовой системы координат с началом в левом верхнем углу кадра и с положительными направлениями из этой точки вниз и вправо. На рис. 3.1 показаны примеры окрестностей различных типов, изображенные в виде совокупностей точек. Центром окрестностей, рабочей точкой, в которой осуществляется обработка, является точка с координатами  $i, j$  (на рис. 3.1 не зачернена). В зависимости от типа окрестности различают *каузальную*, *некаузальную* и *полукаузальную* фильтрацию изображений. Понятие каузальности (причинно-следственной зависимости) связывают с соотношением координат текущей точки  $i, j$  и точек, входящих в окрестность. Если *обе координаты* (номер строки и номер столбца) *всех* точек окрестности не превышают соответствующих координат текущей точки, то окрестность и использующая ее обработка называются каузальными. Пример такой окрестности представлен на рис. 3.1, а.

Некоторые точки окрестности, приведенной на рис. 3.1, б, удовлетворяют принципу каузальности. Вместе с тем здесь имеются и такие точки, *обе координаты* которых превышают соответствующие координаты рабочей точки. Фильтрация, опирающаяся на использование окрестностей с сочетанием таких свойств, называется некаузальной.

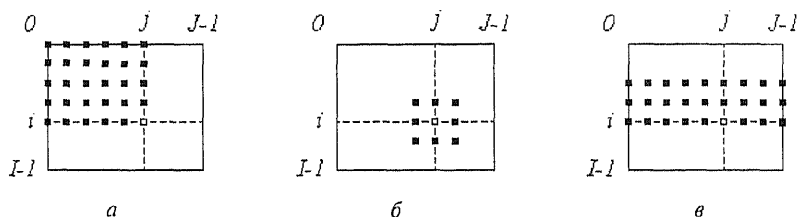


Рис. 3.1. Примеры окрестностей различных видов

Окрестности, показанной на рис. 3.1, в, соответствует полукаузальная фильтрация. Одна из координат *всех* точек окрестности (в данном примере номер строки) не превышает соответствующей координаты рабочей точки. Вторая же координата (в примере номер столбца) у некоторых точек также не превышает соответствующей координаты рабочей точки. Однако среди точек окрестности имеются и такие, у которых эта вторая координата превышает соответствующую координату рабочей точки.

Смысл, заложенный в данную классификацию, состоит в том, что согласно принципу причинности на формирование



отклика физически осуществимого фильтра не могут оказывать влияния элементы входного сигнала, не поступившие к моменту формирования выходного отсчета. Этот принцип естественным образом «работает» в динамических системах, где все происходящие в них процессы являются временными процессами. При цифровой обработке изображений часто приходится иметь дело с ранее сформированными изображениями, уже хранящимися в памяти устройства обработки. В этом смысле соотношение координат, строго говоря, уже не играет такой принципиальной причинной роли, как при обработке сигналов в реальном масштабе времени. Вместе с тем традиционно сложилась описанная выше классификация процедур обработки изображений, которой в определенной мере будем придерживаться и мы в последующем изложении.

При линейной фильтрации выходной эффект определяется линейной комбинацией входных данных:

$$x^*(i, j) = \sum_{(i_1, j_1) \in S} a(i_1, j_1) y(i - i_1, j - j_1). \quad (3.2)$$

В этом выражении  $x^*(i, j) = x_{i,j}^*$  — результат фильтрации (оценка) полезного сигнала  $x(i, j) = x_{i,j}$  в точке кадра с координатами  $(i, j)$ ;  $S$  — множество точек (точнее, множество пар их координат), образующих окрестность;  $a(i_1, j_1)$  — весовые коэффициенты, совокупность которых представляет собой *двумерную импульсную характеристику (ИХ)*. Если область  $S$  конечна, то импульсная характеристика имеет конечную длину и фильтр называется КИХ-фильтром. В противном случае импульсная характеристика имеет бесконечную длину, а фильтр носит название БИХ-фильтра. В выражении (3.2) принято, что ИХ не зависит от координат точки  $(i, j)$ , в которой определяется выходной эффект. Процедуры обработки изображений, обладающие свойством независимости от координат, называются *однородными*.

Наиболее распространенным критерием оптимальности, применяемым для оценки качества обработки, является критерий минимума среднего квадрата ошибок. Применительно к фильтрации запишем его выражение в виде

$$E \left\{ \left[ x(i, j) - \sum_{(i_1, j_1) \in S} a(i_1, j_1) y(i - i_1, j - j_1) \right]^2 \right\} = \min_{a(\cdot)}, \quad (3.3)$$

где  $E\{\circ\}$  — символ математического ожидания. Согласно (3.3) отыскание оптимального фильтра заключается в определении его ИХ таким образом, чтобы средний квадрат ошибки  $\varepsilon(i, j) = x(i, j) - x^*(i, j)$ , выражающей различие между сигналом  $x(i, j)$  и оценкой  $x^*(i, j)$ , формируемой фильтром, был минимальным. Математическое ожидание вычисляется по всем случайным величинам, содержащимся в (3.3), что означает ориентацию критерия на учет *средних* ошибок.

Оптимизационную задачу (3.3) нетрудно свести к решению уравнения или системы уравнений. Для этого вычислим производную от левой части этого выражения по коэффициенту  $a(k, l) = a_{k,l}$  и приравняем ее нулю. Учитывая, что операции дифференцирования, суммирования и вычисления математического ожидания являются линейными и поэтому перестановочны, приходим к выражению

$$E\{x_{i,j}y_{i-k,j-l}\} = \sum_{i_1,j_1 \in S} a_{i_1,j_1} E\{y_{i-i_1,j-j_1}y_{i-k,j-l}\}. \quad (3.4)$$

Входящие в него математические ожидания являются, как нетрудно видеть, отсчетами корреляционных функций, для которых введем следующие обозначения:

$$B_{xy}(k, l) = E\{x_{i,j}y_{i-k,j-l}\}, \\ B_y(k - i_1, l - j_1) = E\{y_{i-i_1,j-j_1}y_{i-k,j-l}\}.$$

С их учетом (3.4) примет более компактный вид:

$$B_{xy}(k, l) = \sum_{i_1,j_1 \in S} a_{i_1,j_1} B_y(k - i_1, l - j_1). \quad (3.5)$$

Считая автокорреляционную  $B_y(\circ)$  и взаимно корреляционную  $B_{xy}(\circ)$  функции известными, замечаем, что (3.5) представляет собой линейное относительно искомым коэффициентов  $a_{i_1,j_1}$  алгебраическое уравнение. Число неизвестных в этом уравнении равняется числу точек  $n_S$  в окрестности  $S$  и является конечным в случае КИХ-фильтра и бесконечным — БИХ-фильтра. Ограничимся в данном разделе рассмотрением КИХ-фильтрации. Линейное алгебраическое уравнение со многими неизвестными имеет бесконечное множество решений. Если по-

вторить дифференцирование (3.3) по остальным  $n_S - 1$  неизвестным, то получим еще  $n_S - 1$  уравнений, отличающихся друг от друга левыми частями  $B_{xy}(\circ)$  и коэффициентами  $B_y(\circ)$  в правых частях, так как определяющие их корреляции вычисляются каждый раз в различных точках. В результате образуется система  $n_S$  линейных алгебраических уравнений с  $n_S$  неизвестными, называемая в теории фильтрации *уравнением Винера-Хопфа*:

$$B_{xy}(k, l) = \sum_{i_1, j_1 \in S} a_{i_1, j_1} B_y(k - i_1, l - j_1), \quad (3.6)$$

$$k, l \in S.$$

Если разрешить ее относительно всех  $n_S$  неизвестных  $a_{i_1, j_1}$ , то будет найдена искомая импульсная характеристика линейного фильтра, минимизирующего средний квадрат ошибок фильтрации.

Определим средний квадрат ошибок оптимальной фильтрации. Для этого необходимо выполнить возведение в квадрат в выражении (3.3) и учесть в полученном выражении уравнение Винера-Хопфа (3.6). В результате нетрудно получить

$$D_e = D_x \left[ 1 - \sum_{i, j \in S} a_{i, j} R_x(i, j) \right], \quad (3.7)$$

где  $D_e = E \{ \varepsilon^2(\circ) \}$  — средний квадрат ошибок фильтрации;  $D_x$ ,  $R_x(\circ)$  — дисперсия и нормированная корреляционная функция выделяемого изображения.

Остановимся на анализе изменения средней яркости изображения при его фильтрации. Вычислив математическое ожидание от обеих частей (3.2), найдем:

$$m_{x^*} = m_y \sum_{(i_1, j_1) \in S} a_{i_1, j_1}, \quad (3.8)$$

где принято, что средняя яркость  $m_y$  входного изображения  $y(\circ)$  не зависит от координат и как результат получено, что и средняя яркость  $m_{x^*}$  выходного изображения  $x^*(\circ)$  также постоянна во всех точках кадра. Очень часто при обработке стремятся сохранить среднюю яркость изображения. Как следует из полученного выражения, достичь этого удастся при выполнении равенства

$$\sum_{(i_1, j_1) \in S} a(i_1, j_1) = 1, \quad (3.9)$$

которое является дополнительным требованием к импульсной характеристике фильтра. Поэтому оптимизационную задачу (3.3) необходимо решать с учетом данного ограничения типа равенства.

Вместо этого часто перед фильтрацией осуществляют вычитание средней яркости  $m_y$  из входного изображения. Как следует из (3.8), среднее значение яркости на выходе фильтра при этом также равно нулю независимо от свойств импульсной характеристики. Это позволяет решать систему уравнений (3.6), игнорируя преобразование средней яркости. Желаемое же ее значение восстанавливается после фильтрации простым прибавлением к выходному эффекту.

### 3.2. МАСОЧНАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ ПРИ НАЛИЧИИ АДДИТИВНОГО БЕЛОГО ШУМА

Распространенным видом помехи является белый шум, аддитивно воздействующий на изображение. Наблюдаемое в этом случае изображение (3.1) имеет вид

$$y_{i,j} = x_{i,j} + n_{i,j}, \quad i = \overline{0, I-1}, j = \overline{0, J-1}, \quad (3.10)$$

а корреляционная функция шума  $n_{i,j}$  описывается выражением

$$B_n(i, j) = D_n \delta_i \delta_j.$$

Здесь  $D_n$  — дисперсия шума, а  $\delta_k = \begin{cases} 1, & \text{при } k = 0 \\ 0, & \text{при } k \neq 0 \end{cases}$  — символ

Кroneкера. Считаем, что входной сигнал центрирован, т.е. имеет нулевое математическое ожидание, а изображение  $x(\circ)$  и шум  $n(\circ)$  взаимно независимы, поэтому для корреляционной функции входного сигнала справедливо

$$B_y(i, j) = D_x R_x(i, j) + D_n \delta_i \delta_j,$$

где  $D_x$  — дисперсия, а  $R_x(i, j)$  — нормированная корреляционная функция полезного сигнала. Нетрудно видеть, что в этих условиях взаимная корреляционная функция  $B_{xy}(\circ)$  совпадает с корреляционной функцией полезного сигнала  $B_x(\circ)$ . Поэтому уравнение Винера — Хопфа (3.6) приводится к виду

$$\left\{ \begin{array}{l} q^2 R_x(k, l) = q^2 \sum_{i_1, j_1 \in S} a_{i_1, j_1} R_x(k - i_1, l - j_1) + a_{k, l} \\ k, l \in S \end{array} \right., \quad (3.11)$$

где  $q^2 = D_x / D_n$  — отношение дисперсий сигнала и шума.

Преобразуем также выражение (3.7) для ошибок фильтрации, для чего запишем в явном виде то из уравнений в (3.11), которое соответствует значениям

$$k = l = 0: q^2 = q^2 \sum_{i_1, j_1 \in S} a_{i_1, j_1} R_x(i_1, j_1) + a_{0,0},$$

откуда находим:

$$a_{0,0}/q^2 = 1 - \sum_{i_1, j_1 \in S} a_{i_1, j_1} R_x(i_1, j_1).$$

Сравнивая это соотношение с (3.7), окончательно получаем:

$$\delta_f^2 = a_{0,0}/q^2,$$

где  $\delta_f^2 = D_f / D_x$  — относительный средний квадрат ошибок фильтрации. Таким образом, для определения ошибок фильтрации необходимо знать отношение сигнал/шум (которое входит также и в уравнение Винера — Хопфа) и значение оптимальной импульсной характеристики в точке (0,0).

Для того чтобы при решении уравнения (3.11) воспользоваться существующими программными средствами ЭВМ, необходимо выполнить его упорядоченное преобразование к каноническому векторно-матричному виду. Для этого требуется совокупность  $n_S$  неизвестных величин  $a_{i,j}$  представить в виде вектора  $A$ . Точно так же множество величин, образующих левые части (3.11), следует представить в виде вектора  $Q$ , а множество коэффициентов правой части в виде матрицы  $F$  размером  $n_S \times n_S$ . Тогда уравнение и его решение примут вид

$$Q = FA, \quad A = F^{-1}Q.$$

В практике цифровой обработки изображений широко используется *масочная фильтрация*. Ее линейная разновидность является одним из вариантов двумерной КИХ-фильтрации. В качестве маски используется множество весовых коэффициентов, заданных во всех точках окрестности  $S$ , обычно симметрично окружающих рабочую точку кадра. Распространенным

видом окрестности, часто применяемым на практике, является квадрат  $3 \times 3$  с рабочим элементом в центре, изображенный на рис. 3.1, б. Применяют различные разновидности масок, одним из эвристических вариантов является равномерная маска, все девять весовых коэффициентов которой равны  $1/9$ . Такой выбор коэффициентов отвечает условию сохранения средней яркости (3.9), и поэтому в процессе обработки центрировать изображение не требуется.

Визуально эффективность фильтрации можно оценить с помощью рис. 3.2. На рис. 3.2, а показан зашумленный портрет (изображение без шума приведено на рис. 2.3, а) при отношении сигнал/шум, равном  $-5$  дБ. Результат масочной фильтрации при оптимальном виде ИХ, найденной из (3.11), приведен на рис. 3.2, б. Результат фильтрации, выполненной равномерным масочным оператором, не приводится, поскольку с визуальной точки зрения он мало отличается от рис. 3.2, б. При этом, однако, с количественной точки зрения различия достаточно заметны: если при оптимальной КИХ относительная ошибка  $\delta_{\text{опт}}^2 = 0.309$ , то при равномерной КИХ она возрастает почти на 30 % и составляет  $\delta_{\text{е равн}}^2 = 0.395$ . Различие резко возрастает при более высоком уровне шума. Так, например, при отношении сигнал/шум, равном  $-10$  дБ, имеем  $\delta_{\text{е опт}}^2 = 0.57$  и  $\delta_{\text{е равн}}^2 = 1.15$ , т.е. применение равномерной КИХ вместо оптимальной приводит в этом случае к увеличению ошибок более чем вдвое.

Здесь полезно отметить определенное разногласие между оценками качества, даваемыми человеческим глазом, и применяемыми количественными показателями. Глаз является слишком совершенным изобретением природы, чтобы с ним могли соревноваться достаточно примитивные математические показатели типа среднего квадрата ошибок. Поэтому некоторые результаты, рассматриваемые с точки зрения математических показателей как катастрофические, визуально могут быть вполне удовлетворительными. Означает ли это, что математические критерии вообще непригодны при цифровой обработке изображений? Конечно, нет. Цифровая обработка изображений применяется в различных информационных системах с автоматическим принятием решений, основанным на этой обработке. Функционирование таких систем, где отсутствует человеческий глаз, полностью подчинено математическим критериям, и качество их работы оценивается только математическими показателями. Понятно, что и качество изображений, используемых в этих



а



б

Рис. 3.2. Пример масочной фильтрации при  $q^2 = -5$  дБ

системах, также должно оцениваться только математическими критериями.

В заключение данного раздела подчеркнем, что в целом применение описанных процедур фильтрации приводит к существенному снижению уровня шума на изображении. Количественно эффективность такой обработки можно охарактеризовать коэффициентом улучшения отношения сигнал/шум  $\gamma = \delta_{\varepsilon}^{-2}/q^2$ , где учтено, что величина  $\delta_{\varepsilon}^{-2}$  определяет отношение сигнал/шум после фильтрации. Улучшение зависит от уровня шума на исходном изображении и составляет в приведенном эксперименте  $\gamma = 10.2$  при  $q^2 = -5$  дБ и  $\gamma = 17.6$  при  $q^2 = -10$  дБ. Коэффициент улучшения тем выше, чем сильнее шум на исходном изображении.

### 3.3. РЕКУРРЕНТНАЯ КАУЗАЛЬНАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ

Проблема борьбы с шумом не решается полностью применением масочных фильтров по следующим причинам. Во-первых, ограниченность размера окрестности, используемой масочным фильтром, приводит к его потенциально ограниченной способности к подавлению шума. Это проявляется при значительном уровне шума на изображении — в меньшей степени при оптимальном выборе КИХ, сильнее при неоптимальной КИХ. Можно, конечно, увеличивать размер окрестности, прибегая к использованию КИХ-фильтров с более длинными импульсными характеристиками. Однако при этом усиливается второй недостаток масочного фильтра, состоящий в его и без того достаточно высокой вычислительной трудоемкости.

В настоящее время отсутствуют методы двумерной фильтрации, в которых сочетаются предельно достижимое качество фильтрации и низкие требования к вычислительным ресурсам ЭВМ, реализующей обработку. Существует много подходов к решению данной проблемы, но все они для достижения компромисса между точностью и реализуемостью прибегают к тем или иным приближениям. Рассмотрим один из них [3.1].

Идея заключается в использовании двумерного БИХ-фильтра с таким *видом* импульсной характеристики, при которой его практическая реализация была бы простой, и с такими *параметрами* этой импульсной характеристики, при которых эффективность фильтрации приближалась бы к потенциально возможной. Создать фильтр с такими свойствами удастся на основе аналогии с одномерным фильтром Калмана.



Наиболее простым примером одномерной фильтрации является калмановская фильтрация однородной стационарной гауссовской последовательности, имеющей корреляционную функцию экспоненциального вида

$$B_x(r) = D_x \exp(-\alpha|r|) = D_x F^{|r|}, \quad F = \exp(-\alpha), \quad \alpha > 0.$$

Здесь  $D_x$  — дисперсия последовательности, а  $F$  — коэффициент ее одношаговой корреляции, определяемый параметром  $\alpha$ , имеющим смысл ширины спектра последовательности. При ее наблюдении на фоне гауссовского белого шума оптимальный каузальный фильтр реализуется рекуррентным алгоритмом, который в стационарном (установившемся) режиме фильтрации имеет вид

$$x_i^* = Fx_{i-1}^* + A(y_i - Fx_{i-1}^*). \quad (3.12)$$

Нетрудно установить, что импульсная характеристика этого фильтра имеет экспоненциальный вид:

$$a_i = A[(1 - A)F]^i, \quad i = 0, 1, 2, \dots, \quad (3.13)$$

где  $A$  — параметр, лежащий в пределах  $0 < A < 1$  и получивший название коэффициента усиления фильтра Калмана. Первое слагаемое в алгоритме (3.12) определяет вклад в оценку сигнала на текущем  $i$ -м шаге фильтрации, вносимый оценкой предыдущего шага, и называется одношаговым прогнозом. Второе учитывает влияние текущего наблюдения  $y_i$  и называется новой информацией. Коэффициент усиления  $A$  определяет чувствительность фильтра к этой новой информации. При высоком уровне шума параметр  $A$  приближается к нулю. При этом, кроме общего снижения ИХ, увеличивается параметр  $B = 1 - A$ , приближаясь к единице. Это означает удлинение импульсной характеристики и, следовательно, сужение полосы пропускаемых фильтром частот. Очевидно, эти свойства ИХ способствуют эффективной фильтрации шума. При снижении шума, наоборот,  $A$  стремится к единице,  $B$  — к нулю, что соответствует расширению полосы частот до бесконечности. Здесь уместно отметить, что фильтрация не только ослабляет действие шума, но, к сожалению, и вносит так называемые динамические искажения в полезный сигнал. Механизм этих искажений очень прост и заключается в неравной передаче на выход фильтра различных спектральных составляющих сигнала. Фильтр Калмана «ведет себя» вполне разумно, когда при исчезновении шума на входе расши-

ряет полосу пропускаемых частот до бесконечности, поскольку именно при этом условии исчезают и динамические ошибки фильтрации.

Отталкиваясь от (3.13) как от одномерного аналога, будем находить двумерную БИХ для каузальной фильтрации изображений от некоррелированного шума в виде двумерной экспоненты:

$$a_{i,j} = A^2(BF)^{i+j}, \quad B = 1 - A, \quad i, j = 0, 1, 2, \dots \quad (3.14)$$

Здесь, как и в случае одномерного фильтра,  $F$  — коэффициент одношаговой корреляции изображения по строке и по столбцу, которые будем здесь считать одинаковыми. Для определения параметра  $A$  (или  $B$ ), остающегося единственным неизвестным параметром двумерного фильтра, воспользуемся уравнением Винера — Хопфа в форме (3.4), переписав его в виде

$$E \left\{ \left( x_{i,j} - \sum_{i_1,j_1 \in S} a_{i_1,j_1} y_{i-i_1,j-j_1} \right) y_{i-k,j-l} \right\} = 0, \quad k, l \in S.$$

Замечая, что выражение в круглых скобках является ошибкой фильтрации, представим эту формулу в виде

$$E \left\{ \varepsilon_{i,j} y_{i-k,j-l} \right\} = 0, \quad k, l \in S. \quad (3.15)$$

Смысл данного выражения состоит в том, что при оптимальной линейной фильтрации ошибка ортогональна всем элементам наблюдаемых данных, используемых при фильтрации. Но тогда нетрудно убедиться и в ортогональности ошибки и результата фильтрации (получаемой оценки)

$$E \left\{ \varepsilon_{i,j} x_{i,j}^* \right\} = 0, \quad (3.16)$$

для чего достаточно вычислить левую часть этого выражения с учетом (3.2) и (3.15).

Для дальнейшего необходимо воспользоваться в (3.16) принятым представлением импульсной характеристики (3.14), в результате данное соотношение превращается в уравнение относительно искомого параметра  $A$ . В него входят корреляционная функция сигнала и взаимно корреляционная функция сигнала и наблюдаемых данных. Поэтому необходимо конкретизировать корреляционную функцию сигнала, в качестве которой воспользуемся биэкспоненциальным представлением:

$$B_x(i, j) = D_x F^{|i|+|j|}. \quad (3.17)$$

С учетом этого, считая, что кадр имеет бесконечные размеры (это позволяет принять бесконечными соответствующие пределы суммирования в (3.2)), можно получить следующее алгебраическое уравнение:

$$1 - \left[ 1 - \frac{B(1 - F^2)}{1 - B^2 F^2} \right]^2 - \frac{(1 - B)^2}{q^2} \frac{(1 - BF^2)^2}{(1 - B^2 F^2)^2} = 0 \quad (3.18)$$

относительно параметра  $B$ , численное решение которого не представляет проблемы. Анализируя (3.18), легко заметить, что при  $q^2 \rightarrow \infty$  это уравнение удовлетворяется при  $B = 0$ , а при  $q^2 \rightarrow 0$  его решением является  $B = 1$ . Эти предельные значения можно интерпретировать как изменения частотных характеристик двумерного фильтра, аналогичные поведению параметров фильтра Калмана в подобных предельных ситуациях.

Подставив в (3.7) выражения ИХ (3.14) и корреляционной функции (3.17), можно получить следующую формулу для среднего квадрата ошибок фильтрации:

$$D_e = D_x \left[ 1 - \frac{A^2}{(1 - BF^2)^2} \right].$$

Подставив далее выражение ИХ (3.14) в (3.2), можно привести выражение отклика фильтра к виду

$$x_{i,j}^* = BF (x_{i-1,j}^* + x_{i,j-1}^*) - (BF)^2 x_{i-1,j-1}^* + Ay_{i,j}. \quad (3.19)$$

Рекуррентный характер алгоритма (3.19) является важным положительным качеством рассматриваемого фильтра. Как следует из (3.19), его работа требует выполнения на каждом шаге обработки всего трех операций умножения и трех сложений, причем структура алгоритма универсальна и, в частности, не зависит от отношения сигнал/шум. Для сравнения, масочный фильтр с размером окрестности  $3 \times 3$  элементов требует выполнения на каждом шаге при общем виде КИХ девяти умножений и восьми сложений. Таким образом, по количеству операций рекуррентный фильтр выигрывает у простейшего масочного практически в три раза. Очевидно, что попытка улучшить качество фильтрации масочным фильтром за счет увеличения размера применяемой окрестности приводит к увеличению числа операций и дальнейшему увеличению его проигрыша по этой характеристике.

При фильтрации реальных изображений ограниченного размера возникает граничная проблема получения оценок в точках нулевой строки и нулевого столбца. Естественным решением является использование здесь обычной (одномерной) калмановской фильтрации. Пример применения описанного двумерного фильтра показан на рис. 3.3, где представлен результат эксперимента с тем же портретом и при том же отношении сигнал/шум  $-5$  дБ, что и при испытании масочного фильтра. Поэтому здесь не приводится показанное на рис. 3.2, *а* входное изображение с шумом. Результат двумерной рекуррентной фильтрации представлен на рис. 3.3, *а*, а на рис. 3.3, *б* для сравнения повторен результат оптимальной масочной фильтрации (рис. 3.2, *б*). Визуальная оценка говорит в пользу двумерного рекуррентного фильтра, поскольку уровень остаточного шума на рис. 3.3, *а* ниже. Сравнение по среднему квадрату ошибок совпадает с субъективной оценкой: величина  $\delta_e^2$  при масочной фильтрации составляет, как говорилось ранее,  $0.309$ , а при двумерной рекуррентной —  $0.29$ . Различие заметно усиливается при более высоком уровне шума. Так, при отношении сигнал/шум  $-10$  дБ имеем соответственно  $\delta_e^2$  равную  $0.57$  и  $0.43$ .

Необходимо отметить, однако, следующее. Вместе с уменьшением уровня шума при двумерной рекуррентной фильтрации наблюдается более значительная (по сравнению с масочным фильтром) утрата резкости обработанного изображения. Это является результатом упоминавшихся выше динамических искажений, более сильных при бесконечной импульсной характеристике, чем при конечной.

Кроме того, рассмотренный двумерный фильтр не абсолютно оптимален, поскольку его структура определена волевым решением при выборе ИХ в виде (3.14). Поэтому и получаемое при его помощи ослабление шума не является предельным.

### 3.4. ПРИМЕНЕНИЕ ФИЛЬТРА ВИНЕРА ДЛЯ НЕКАУЗАЛЬНОЙ ДВУМЕРНОЙ ФИЛЬТРАЦИИ

Потенциально наилучшие результаты обработки изображения, в частности результаты фильтрации, достигаются при применении некаузального принципа, поскольку этот принцип основан на использовании абсолютно всех исходных данных при обработке каждой точки кадра. Понятно, что при рациональном использовании этих данных получаемый эффект максимален. Одним из известных вариантов линейной некаузальной филь-



*a*



*б*

Рис. 3.3. Пример двумерной рекуррентной фильтрации

рации изображений является винеровская фильтрация. Ее применение связано с предположением о стационарности изображения. Поскольку наличие краев изображения приводит к нарушению стационарности, винеровская фильтрация не строго оптимальна. Однако при размерах кадра, значительно превышающих интервал корреляции изображения, влияние границ мало. Эти соображения служат важным стимулом к применению винеровской фильтрации для борьбы с шумом.

Технически фильтр Винера реализуется при помощи дискретного преобразования Фурье *в частотной области*. Поэтому, прежде чем рассматривать уравнение Винера-Хопфа, которое остается методологической основой фильтрации помех и в этом случае, нам необходимо познакомиться с двумерным дискретным преобразованием Фурье (ДПФ), некоторыми его свойствами и принципами линейной фильтрации на основе ДПФ.

#### 3.4.1. ДВУМЕРНОЕ ДИСКРЕТНОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Обозначим через

$$x(i_1, i_2) = x_{i_1, i_2}, \quad i_1 = \overline{0, I_1 - 1}, \quad i_2 = \overline{0, I_2 - 1} \quad (3.20)$$

двумерное поле (двумерный сигнал), описывающее дискретное изображение размера  $I_1$  строк и  $I_2$  столбцов. Вне указанных границ этот сигнал не задан. Выполним *периодическое продолжение* данного финитного сигнала, введя двумерный периодический сигнал

$$\tilde{x}_{i_1, i_2} = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} x_{i_1 - m_1 I_1, i_2 - m_2 I_2}. \quad (3.21)$$

Если сигнал  $x_{i_1, i_2}$  существует только внутри прямоугольника

$$R_{I_1, I_2} = \{(i_1, i_2) : 0 \leq i_1 \leq I_1 - 1, 0 \leq i_2 \leq I_2 - 1\}$$

со сторонами  $I_1 \times I_2$  элементов (рис. 3.4, а), то сигнал  $\tilde{x}_{i_1, i_2}$  определен на всей плоскости  $(i_1, i_2)$  и является на ней *прямоугольно-периодическим* (рис. 3.4, б).

Любой периодический сигнал может быть представлен в виде *ряда Фурье*, но, в отличие от одномерных сигналов, двумерные описываются двумерным рядом Фурье, имеющим вид:

$$\tilde{x}_{i_1, i_2} = \frac{1}{I_1 I_2} \sum_{k_1=0}^{I_1-1} \sum_{k_2=0}^{I_2-1} \tilde{X}_{k_1, k_2} \exp \left( j \frac{2\pi}{I_1} i_1 k_1 + j \frac{2\pi}{I_2} i_2 k_2 \right), \quad j = \sqrt{-1}. \quad (3.22)$$

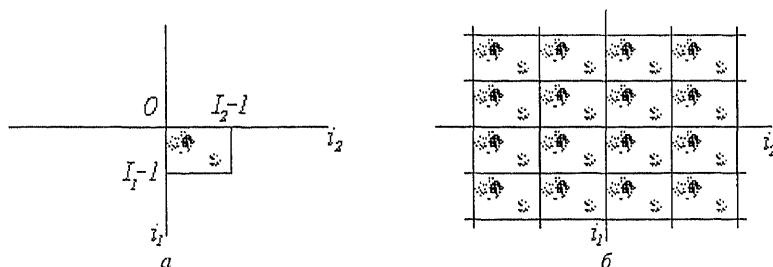


Рис. 3.4. Реальное (а) и периодически продолженное (б) изображения

Базисные функции этого двумерного представления — двумерные комплексные экспоненты (иногда называемые комплексными синусоидами)

$$\varphi_{k_1, k_2}(i_1, i_2) = \frac{1}{I_1 I_2} W_1^{i_1 k_1} W_2^{i_2 k_2}, \quad (3.23)$$

$$W_1 = \exp\left(j \frac{2\pi}{I_1}\right), \quad W_2 = \exp\left(j \frac{2\pi}{I_2}\right),$$

имеющие, как и сигнал  $\tilde{x}(\cdot)$ , прямоугольную периодичность с тем же периодом  $I_1 \times I_2$ . Здесь  $(k_1, k_2)$  — двумерный номер базисной функции, а величины  $\frac{2\pi k_{1(2)}}{I_{1(2)}}$  имеют смысл пространственных частот. Иногда пространственными частотами называют целочисленные величины  $k_1$  и  $k_2$ .

Коэффициенты Фурье  $\tilde{X}_{k_1, k_2}$  ряда (3.22) образуют *двумерный частотный спектр* сигнала  $\tilde{x}_{i_1, i_2}$  и определяются формулой *прямого преобразования Фурье*

$$\tilde{X}_{k_1, k_2} = \sum_{i_1=0}^{I_1-1} \sum_{i_2=0}^{I_2-1} \tilde{x}_{i_1, i_2} W_1^{-i_1 k_1} W_2^{-i_2 k_2}. \quad (3.24)$$

Выражение (3.22), восстанавливающее сигнал  $\tilde{x}_{i_1, i_2}$  по его спектру  $\tilde{X}_{k_1, k_2}$ , является *обратным преобразованием Фурье*. В справедливости преобразований (3.22) и (3.24), называемых *двумерным ДПФ*, можно убедиться, подставив (3.24) в (3.22) и приведя правую часть полученного равенства к значению левой, т.е. к  $\tilde{x}_{i_1, i_2}$ .

Заметим, что для *точного представления* дискретного сигнала  $\tilde{x}_{i_1, i_2}$  с двумерным периодом  $I_1 \times I_2$  элементов согласно формулам БПФ достаточно *конечного* числа базисных функций (3.23), поэтому ряд (3.22) является *конечным*. Это и понятно, поскольку сам представляемый сигнал содержит в одном периоде конечное число точек, т.е. имеет конечное число степеней свободы. Ясно, что число степеней свободы в спектре не может отличаться от числа степеней свободы в самом сигнале.

Остановимся на наиболее существенных свойствах двумерного дискретного спектра Фурье. Вычислим спектральные коэффициенты (3.24) в частотных точках  $k_1 + m_1 I_1$ ,  $k_2 + m_2 I_2$ :

$$\begin{aligned} \tilde{X}_{k_1 + m_1 I_1, k_2 + m_2 I_2} &= \sum_{i_1=0}^{I_1-1} \sum_{i_2=0}^{I_2-1} \tilde{x}_{i_1, i_2} W_1^{-i_1 k_1} W_2^{-i_2 k_2} \times \\ &\times \exp(-j \cdot 2\pi i_1 m_1 - j \cdot 2\pi i_2 m_2) . \end{aligned} \quad (3.25)$$

Поскольку при любых целых значениях  $m_1$  и  $m_2$  последний множитель в полученном выражении равен единице, имеем отсюда равенство

$$\tilde{X}_{k_1 + m_1 I_1, k_2 + m_2 I_2} = \tilde{X}_{k_1, k_2} ,$$

означающее *прямоугольную периодичность двумерного ДПФ*. Следовательно, картина двумерного ДПФ подобна картине двумерного периодически продолженного сигнала, качественно показанной на рис. 3.4, б (если на ней пространственные координаты  $i_1$ ,  $i_2$  заменить частотными  $k_1$ ,  $k_2$ ). Однако необходимо иметь в виду, что спектральные коэффициенты  $\tilde{X}_{k_1, k_2}$ , как это следует из (3.24), являются *комплексными числами*, в том числе и при вещественном сигнале  $\tilde{x}_{i_1, i_2}$ . Но тогда возникает вопрос, почему для описания спектра требуется удвоенное число вещественных чисел. Общее количество спектральных составляющих, как установлено, равно  $I_1 I_2$ . Комплексное число эквивалентно паре вещественных чисел — действительной и мнимой частям при алгебраическом или модулю и фазе при экспоненциальном представлении. Следовательно, полный спектр описывается  $2I_1 I_2$  вещественными числами, что вдвое превышает размерность самого сигнала  $\tilde{x}_{i_1, i_2}$ . В этом, на первый взгляд, содержится противоречие. Оно находит свое разъяснение при дальнейшем изучении свойств двумерного ДПФ.



Преобразуем соотношение (3.25) следующим образом. Во-первых, вместо частот  $k_1, k_2$  подставим частоты  $-k_1, -k_2$ . Во-вторых, выполним комплексное сопряжение обеих частей, что не нарушит равенства. В результате нетрудно получить выражение

$$\tilde{X}_{k_1, k_2} = \tilde{X}_{m_1 I_1 - k_1, m_2 I_2 - k_2}^*,$$

которым устанавливается однозначная связь между спектральными коэффициентами в двух различных точках спектрального прямоугольника  $R_{I_1, I_2} = \{(k_1, k_2) : 0 \leq k_1 \leq I_1 - 1, 0 \leq k_2 \leq I_2 - 1\}$ .

Полученным соотношением и снимается противоречие, поскольку количество независимых спектральных коэффициентов уменьшается благодаря данной спектральной симметрии в два раза. Согласно установленному свойству спектрально-сопряженной зависимости связаны между собой спектральные коэффициенты, принадлежащие левому верхнему и правому нижнему углам прямоугольника  $R_{I_1, I_2}$ . Подобным образом связаны между собой коэффициенты Фурье из правого верхнего и левого нижнего участков спектрального прямоугольника  $R_{I_1, I_2}$ .

В заключение данного пункта укажем, что при практическом применении двумерного ДПФ — как прямого, так и обратного, совсем не обязательно оперировать периодическими сигналами и спектрами, как это предполагается, казалось бы, преобразованиями (3.22) и (3.24). От этой необходимости избавляют сами соотношения (3.22) и (3.24). В самом деле, прямое преобразование Фурье (3.24) содержит в правой части значения периодически продолженного сигнала  $\tilde{x}_{i_1, i_2}$  лишь в пределах одного «главного» прямоугольника  $R_{I_1, I_2}$ . Но в этих пределах исходный  $x_{i_1, i_2}$  и периодически продолженный  $\tilde{x}_{i_1, i_2}$  сигналы полностью совпадают, что дает возможность использовать в формуле (3.24) исходный сигнал  $x_{i_1, i_2}$ . Аналогичные пояснения можно сделать и относительно обратного преобразования (3.22), откуда следует, что практически в процессе вычислений оперировать надо «основным» участком спектра, относящимся к спектральной области  $R_{I_1, I_2}$ .

Из сделанных пояснений, имеющих лишь исключительно вычислительное значение, не следует делать вывода об искусственности и ненужности рассмотренных математических моделей периодических полей. При обработке изображений возникают многочисленные задачи, правильное толкование и решение ко-

торых возможно только на основе этих математических интерпретаций. Одной из таких важнейших задач является цифровая двумерная фильтрация в спектральной области, осуществление которой связано с выполнением так называемой *циклической свертки*.

### 3.4.2. ЦИКЛИЧЕСКАЯ СВЕРТКА

Достоинство обычного преобразования Фурье в том, что с его помощью очень просто выполняется стационарная фильтрация сигналов. Как известно, для этого требуется получить частотный спектр сигнала и частотный коэффициент передачи фильтра. Затем, перемножив их, определить частотный спектр выходного сигнала, а выполнив обратное преобразование Фурье, — найти сам выходной сигнал. Возможно ли применение такой технологии для выполнения двумерной цифровой фильтрации? Убедимся, что возможно, но с некоторыми оговорками.

Двумерный стационарный (инвариантный к сдвигу) фильтр задают двумерной импульсной характеристикой  $a_{i_1, i_2}$ , позволяющей определить выходной сигнал при помощи двумерной свертки:

$$y_{i_1, i_2} = \sum_{r_1=0}^{I_1-1} \sum_{r_2=0}^{I_2-1} x_{r_1, r_2} a_{i_1-r_1, i_2-r_2}. \quad (3.26)$$

Это уравнение *обычной* свертки, в нем фигурируют обычные, непериодические сигналы и непериодическая ИХ фильтра. При цифровой обработке в ЭВМ не существует частотных спектров, соответствующих таким сигналам и ИХ. Для описания сигнала в частотной области привлекается, как установлено выше, периодически продолженный сигнал  $\tilde{x}_{i_1, i_2}$ , которому соответствует дискретный спектр  $\tilde{X}_{k_1, k_2}$ . По аналогии с (3.21) вводится и периодически продолженная ИХ:

$$\tilde{a}_{i_1, i_2} = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} a_{i_1-m_1, i_2-m_2},$$

двумерное ДПФ которой  $\tilde{A}_{k_1, k_2}$  имеет смысл частотного коэффициента передачи цифрового фильтра. Перемножая  $\tilde{X}_{k_1, k_2}$  и  $\tilde{A}_{k_1, k_2}$ , мы находим спектр сигнала, определяемого *циклической сверткой*. Циклическая свертка отличается от обычной свертки

(3.26) тем, что вместо функций  $x_{i_1, i_2}$  и  $a_{i_1, i_2}$  в ней представлены периодически продолженные функции  $\tilde{x}_{i_1, i_2}$  и  $\tilde{a}_{i_1, i_2}$ . Нетрудно установить, что при этом сигнал на выходе

$$\tilde{y}_{i_1, i_2} = \sum_{r_1=0}^{I_1-1} \sum_{r_2=0}^{I_2-1} \tilde{x}_{r_1, r_2} \tilde{a}_{i_1-r_1, i_2-r_2} \quad (3.27)$$

также является периодическим. Покажем, что именно его спектр  $\tilde{Y}_{k_1, k_2}$  определяется выражением

$$\tilde{Y}_{k_1, k_2} = \tilde{X}_{k_1, k_2} \tilde{A}_{k_1, k_2}. \quad (3.28)$$

Умножим для этого левую и правую части (3.27) на  $W_1^{-i_1 k_1} W_2^{-i_2 k_2}$  и просуммируем по  $i_1$  и  $i_2$  в пределах области  $R_{I_1, I_2}$ . В левой части в результате имеем спектр  $\tilde{Y}_{k_1, k_2}$ . Преобразуем правую часть, предварительно умножив ее на величину  $W_1^{r_1 k_1} W_2^{r_2 k_2} W_1^{-r_1 k_1} W_2^{-r_2 k_2}$ , тождественно равную единице:

$$\tilde{Y}_{k_1, k_2} = \sum_{r_1=0}^{I_1-1} \sum_{r_2=0}^{I_2-1} \tilde{x}_{r_1, r_2} W_1^{-r_1 k_1} W_2^{-r_2 k_2} \sum_{r_1=0}^{I_1-1} \sum_{r_2=0}^{I_2-1} \tilde{a}_{i_1-r_1, i_2-r_2} W_1^{-(i_1-r_1)k_1} W_2^{-(i_2-r_2)k_2}.$$

Выполняя затем во внутренней сумме замену переменных  $m_1 = i_1 - r_1$ ,  $m_2 = i_2 - r_2$ , учитывая, что при суммировании периодической функции в пределах периода границы суммирования могут быть синхронно сдвинуты на любую величину, и оставляя их в этой связи неизменными, приходим к (3.28).

Таким образом, установлено, что именно циклической свертке (3.27) соответствует удобный при практической реализации частотный метод определения выходного сигнала. Необходимо иметь в виду, что та фильтрация, к осуществлению которой обычно стремятся при решении конкретных задач, описывается обычной сверткой (3.26), а не циклической (3.27). Поэтому остается выяснить, в каком соотношении находятся эти две вычислительные процедуры. Не нарушая общности, для большей простоты и наглядности рассмотрим этот вопрос на примере одномерной фильтрации. Сигнал на выходе одномерного фильтра определяется сверткой

$$y_i = \sum_{r=0}^{I-1} x_r a_{i-r}, \quad (3.29)$$

$$\tilde{y}_i = \sum_{r=0}^{I-1} \tilde{x}_r \tilde{a}_{i-r} \quad (3.30)$$

в обычном и циклическом вариантах соответственно. Рис. 3.5 поясняет процесс вычисления выходного сигнала в обоих случаях. На рис. 3.5, *а* показана импульсная характеристика произвольного вида, соответствующая некаузальному фильтру (так как  $a(j) \neq 0$  при  $j < 0$ ). Рис. 3.5, *б* иллюстрирует образование суммы, вычисляемой при помощи обычной свертки (3.29), а рис. 3.5, *в* при помощи циклической (3.30). Области суммирования, выполняемого в соответствии с выражениями (3.29) и (3.30), заштрихованы. Рисунки отражают определение реакции фильтра в точке  $i$ , расположенной вблизи границы рабочей области. В случае циклической свертки область суммирования является двухсвязной из-за периодичности ИХ, что приводит к различию результатов фильтрации. Очевидно, что эффекты, вызванные периодичностью, отсутствуют для точек, удаленных от границ. Поэтому для внутренних точек области  $R_I = \{i : 0 \leq i \leq I-1\}$ , удаленность которых от границ превышает длину импульсной характеристики, результаты обычной и циклической свертки совпадают. Различия наблюдаются лишь для точек, примыкающих к границе. Если размеры этой приграничной области относительно невелики, то часто различиями пренебрегают. В тех случаях, когда граничные эффекты недопустимы, проблема может быть разрешена при помощи искусственного удлинения области  $R_I$  добавлением к ней на обоих концах такого количества нулевых элементов, при котором эффект цикличности проявляться не будет.

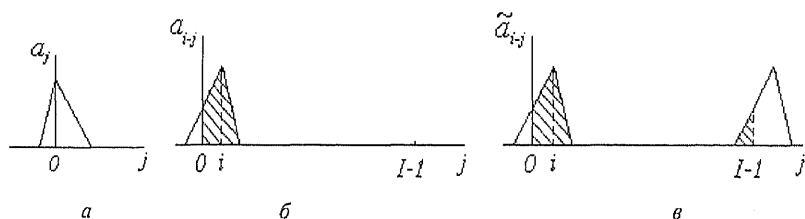


Рис. 3.5. Сравнение обычной и циклической свертки

В заключение отметим, что характер различий обычной и циклической свертки при фильтрации двумерных цифровых сигналов остается таким же, как для одномерных сигналов. Аналогичны и меры, исключаяющие нежелательные эффекты циклической свертки.

### 3.4.3. РЕШЕНИЕ УРАВНЕНИЯ ВИНЕРА-ХОПФА В ЦИКЛИЧЕСКОМ ПРИБЛИЖЕНИИ

Вернемся к задаче некаузальной фильтрации шума на изображении. Оптимальный линейный фильтр определяется и в этом случае уравнением Винера-Хопфа (3.6), в котором для начала область существования  $S$  импульсной характеристики принимаем неограниченной. В результате имеем:

$$\begin{cases} B_{xy}(k, l) = \sum_{i_1=-\infty}^{\infty} \sum_{j_1=-\infty}^{\infty} a_{i_1, j_1} B_y(k - i_1, l - j_1), \\ k, l \in (-\infty, \infty) \end{cases} \quad (3.31)$$

Дискретный винеровский фильтр удастся легко найти в циклическом приближении. Для этого надо вместо реальных функций  $B_{xy}(\cdot)$ ,  $B_y(\cdot)$  и  $a_{i_1, j_1}$  подставить в уравнение (3.31) соответствующие периодически продолженные функции  $\tilde{B}_{xy}(\cdot)$ ,  $\tilde{B}_y(\cdot)$  и  $\tilde{a}_{i_1, j_1}$ , имеющие двумерный период  $R_{I_1, I_2}$ . При этом область определения ИХ также сужается до размеров прямоугольника  $R_{I_1, I_2}$ . Поэтому (3.31) принимает вид

$$\begin{cases} \tilde{B}_{xy}(k, l) = \sum_{i_1=0}^{I_1-1} \sum_{j_1=0}^{I_2-1} \tilde{a}_{i_1, j_1} \tilde{B}_y(k - i_1, l - j_1), \\ k, l \in R_{I_1, I_2} \end{cases} \quad (3.32)$$

Периодичность функций, входящих в уравнение (3.32), позволяет применить к его обеим частям двумерное ДПФ, подобно тому как это было сделано выше применительно к уравнению (3.27). В результате получаем:

$$\tilde{G}_{x,y}(k_1, k_2) = \tilde{A}_{k_1, k_2} \tilde{G}_y(k_1, k_2). \quad (3.33)$$

В этом выражении  $\tilde{G}_{x,y}(k_1, k_2)$ ,  $\tilde{G}_y(k_1, k_2)$  — спектральные плотности мощности, представляющие собой ДПФ соответствующих корреляционных функций, а  $\tilde{A}_{k_1, k_2}$  — ДПФ искомой импульсной характеристики, являющееся, таким образом, частотным коэффициентом передачи фильтра Винера. Все функции, входящие в (3.33), являются прямоугольно-периодическими с двумерным периодом  $R_{I_1, I_2}$ . Основное достижение, вызванное применением ДПФ, состоит в преобразовании сложного для

решения уравнения Винера — Хопфа (3.32) в простейшее алгебраическое уравнение (3.33), решение которого, правда, не для импульсной, а для частотной характеристики имеет вид:

$$\tilde{A}_{k_1, k_2} = \frac{\tilde{G}_{x,y}(k_1, k_2)}{\tilde{G}_y(k_1, k_2)}. \quad (3.34)$$

Найденное решение дает эффективный способ осуществления оптимальной линейной фильтрации изображения. Его реализация требует знания спектральных плотностей мощности и опирается на применение к обрабатываемому изображению дискретного преобразования Фурье.

Не следует, однако, забывать, что переход от уравнения (3.31), определяющего истинно оптимальную характеристику фильтра, к уравнению (3.32), позволяющему найти ее периодически продолженный аналог, был выполнен без достаточного обоснования. Поэтому ничего нельзя пока сказать о том, в какой степени найденное решение близко к истинно оптимальному. Для ответа на этот вопрос рассмотрим снова для простоты одномерные аналоги уравнений (3.31) и (3.32), имеющие вид

$$\begin{cases} B_{xy}(k) = \sum_{i=-\infty}^{\infty} a_i B_y(k-i), \\ k \in (-\infty, \infty) \end{cases} \quad (3.35)$$

$$\begin{cases} \tilde{B}_{xy}(k) = \sum_{i=0}^I \tilde{a}_i \tilde{B}_y(k-i), \\ k \in [0, I-1] \end{cases} \quad (3.36)$$

Рис. 3.6 иллюстрирует формирование сумм, входящих в правые части этих равенств при некотором произвольном значении сдвига  $k$  и достаточно большом значении интервала наблюдения  $I$ . Здесь показаны некаузальная ИХ и корреляционная функция входного сигнала, а штриховкой условно отмечена область суммирования. Из сравнения рис. 3.6, а, соответствующего (3.35), и рис. 3.6, б, соответствующего (3.36), видно, что, хотя во втором случае область суммирования и является двухсвязной, это не приводит к различию результатов суммирования. При большом значении длины интервала  $I$  соседние зоны на рис. 3.6, б не перекрываются, благодаря чему искажения результатов не происходит. Следовательно, уравнения (3.35) и (3.36) являются эквивалентными.

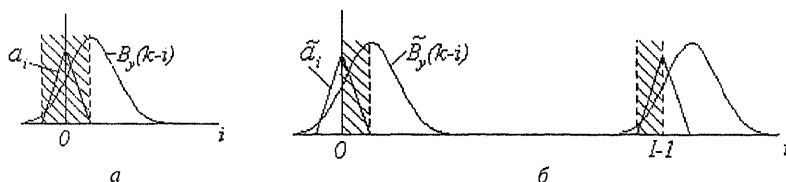


Рис. 3.6. Сравнение обычного и циклического уравнений Винера-Хопфа

Если же интервал  $I$  будет соизмерим с размахом корреляционной функции  $B_y(\cdot)$ , то произойдет наложение соседних областей периодической картины, что в итоге приведет к изменению значений функций, стоящих под знаком суммы в (3.36), и исказит уравнение Винера — Хопфа. Таким образом, условие применимости фильтра Винера, определяемого соотношением (3.34), состоит в его использовании для обработки изображений, имеющих достаточно большие размеры. Напомним также, что в данном пособии обсуждается уравнение Винера — Хопфа для стационарных сигналов и изображений. Поэтому вблизи границ обрабатываемого кадра, где само их существование приводит к нарушению этого условия, обработка отклоняется от оптимальной.

На рис. 3.7 приведен пример работы фильтра Винера. Как и ранее, эксперимент выполнен при отношении сигнал/шум  $q^2 = -5$  дБ. Относительный средний квадрат ошибок фильтрации  $\delta_e^2$  составляет в этом эксперименте величину 0,167, что является наилучшим показателем среди всех рассмотренных методов линейной фильтрации (напомним, что при масочной фильтрации выше было получено  $\delta_e^2 = 0.309$ , а при рекуррентной  $\delta_e^2 = 0.29$ ).

О наименьшем уровне остаточного шума на изображении говорит и визуальная оценка результата. Хотя нельзя не отметить, что это достигается ценой большей, чем при других методах, дефокусировки изображения. В этом проявляется общее диалектическое противоречие между борьбой с помехами и динамическими искажениями обрабатываемого изображения, свойственное, как отмечалось и ранее, всем методам фильтрации.

Обработка изображений при помощи фильтра Винера требует использования спектральной плотности мощности изображения. Существуют различные способы получения нужной информации. Один из них основан на предварительном измерении необходимых характеристик по реальному изображению. Полученные

*а**б*

Рис. 3.7. Пример винеровской фильтрации шума  
при  $q^2 = -5$  дБ



при этом спектральные плотности вводятся в ЭВМ в виде таблиц, позволяя задать коэффициент передачи в численном виде. Другой способ, примененный в представленном эксперименте, состоит в использовании некоторой математической модели изображения, вид спектрально-корреляционных характеристик которой известен. В этом случае отдельные параметры, входящие в используемую математическую модель, измеряются по реальному изображению. При проведении эксперимента, описанного выше, в частности, применялась модель изображения в виде гауссовского двумерного поля с корреляционной функцией (3.17), а измерялись коэффициент одношаговой корреляции  $F$  и дисперсия  $D_x$ .

Анализ метода будет неполным, если не сделать оценки вычислительной эффективности реализующей его процедуры. Для ДПФ разработаны эффективные вычислительные методы, воплощенные в процедурах быстрого преобразования Фурье (БПФ). Количество комплексных умножений, составляющих основную трудоемкость двумерного БПФ, оценивают числом  $I_1 I_2 \log_2 I_1 I_2$  [3.2]. Поскольку полный цикл обработки предполагает выполнение прямого и обратного БПФ, это число следует удвоить. По отношению к одному элементу кадра число умножений, таким образом, составляет  $2 \log_2 I_1 I_2$ . При  $I_1 = I_2 = 256$  число умножений в каждой точке кадра равно 32. Для сравнения напомним, что, например, рекуррентный двумерный фильтр, описанный выше, реализуется всего тремя вещественными умножениями в каждой точке кадра (при различных значениях одношагового коэффициента корреляции изображения по строкам и по столбцам — четырьмя умножениями).

### 3.5. БАЙЕСОВСКАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ

При всех рассмотренных ранее методах фильтрации с самого начала предполагалось отыскание фильтра в классе линейных систем. Отсюда следует, что могут существовать нелинейные процедуры, обладающие лучшими качественными характеристиками, чем приведенные выше. Для их отыскания необходим более общий подход к фильтрации, чем тот, который опирается на решение уравнение Винера — Хопфа. Общепринятая достаточно универсальная идеология фильтрации использует байесовский принцип. Ее применение позволяет, по крайней мере теоретически, создавать как линейные, так и нелинейные алгоритмы фильтрации. Кроме того, этот принцип помогает выяснить, при

каких условиях линейные процедуры фильтрации приводят к наивысшему качеству обработки и, следовательно, являются абсолютно оптимальными.

Отметим, однако, с самого начала основные недостатки байесовской фильтрации изображений. Первый является общим для байесовских методов вообще и заключается в очень высоких требованиях к объему и характеру данных, содержащихся в математических моделях сигналов и помех, удовлетворить которым на практике удастся далеко не всегда. Второй связан со спецификой изображений как двумерных сигналов, что приводит к колоссальным вычислительным трудностям при попытке прямого использования этого подхода. Последнее делает актуальной разработку методов, способных преодолеть данное ограничение. Достичь этого обычно удастся ценой определенной потери качества обработки. Поэтому дело сводится к поиску таких методов, при которых потеря качества была бы приемлемой.

### 3.5.1. СУЩНОСТЬ БАЙЕСОВСКОЙ ФИЛЬТРАЦИИ

Полагаем, что на входе фильтра действует сигнал

$$y_{i_1, i_2} = f(x_{i_1, i_2}, n_{i_1, i_2}), \quad i_1 = \overline{0, I_1 - 1}, \quad i_2 = \overline{0, I_2 - 1}, \quad (3.37)$$

где  $x_{i_1, i_2}$  и  $n_{i_1, i_2}$  — полезный сигнал и помеха, а  $f(\cdot)$  — функция, описывающая их взаимодействие. При байесовском методе считается, что сигнал и помеха — случайные процессы (случайные двумерные поля) с известными законами распределения вероятностей. Пусть  $X$  — вектор, элементы которого — все  $I_1 I_2$  отсчетов, образующих кадр изображения, а  $w_x(X)$  — их совместное распределение. Примем для простоты, что помеха и сигнал независимы, а распределение вектора помехи  $N$  равно  $w_n(N)$ . Воспользовавшись формулой Байеса, запишем *апостериорное распределение вероятностей* (АРВ)  $w_{x|y}(X|Y)$ :

$$w_{x|y}(X|Y) = \frac{w_x(X)w_{y|x}(Y|X)}{w_y(Y)}, \quad (3.38)$$

куда входят распределение  $w_y(Y)$  наблюдаемых данных и условное распределение  $w_{y|x}(Y|X)$ , называемое функцией правдоподобия. Смысл выражения (3.38) заключается в том, что оно

дает возможность вычислить в устройстве обработки распределение вероятностей полезного сигнала, располагая входными данными  $Y$  и опираясь на вероятностную модель как самого полезного сигнала, так и наблюдаемых данных. АРВ является аккумулятором всех доступных сведений о полезном сигнале, которые содержатся в  $Y$ , а (3.38) указывает способ извлечения этих сведений.

Поскольку потребителя информации обычно интересует точечное значение сигнала  $X$ , для его образования прибегают к вычислению либо математического ожидания АРВ, либо его координаты, обращающей это распределение в максимум [3.3].

Оперировать векторными величинами, входящими в (3.38), практически невозможно из-за громадной размерности векторов  $X$  и  $Y$ . Если, например, обрабатываемый кадр имеет размеры  $I_1 = I_2 = 256$ , то размерность этих векторов равна  $256^2 = 65536$ . Предположим, что изображение является простейшим с бинарными значениями элементов  $x_0$  и  $x_1$ . Общее число всевозможных изображений, имеющих всего две градации яркости, составляет  $2^{65536} > 10^{19000}$ . Задача байесовской фильтрации – вычисление распределения вероятностей  $w_{x|y}(X|Y)$ , которое можно представить в данном случае в виде таблицы с размером, превышающим  $10^{19000}$ . Явная нереальность такой задачи заставляет искать методы описания сигналов, которые приводили бы к резкому, качественному ее упрощению. В этом направлении предпринимаются усилия, разрабатываются различные подходы [3.4 – 3.6], но, к сожалению, универсальных эффективных методов двумерной байесовской обработки изображений, основанных на использовании всех данных  $Y$ , в настоящее время не найдено.

Отмеченная сложность байесовских процедур свойственна и фильтрации одномерных сигналов. Вместе с тем в области одномерной фильтрации были получены блестящие решения проблемы, основанные на использовании марковских моделей сигналов и помех. В указанных работах [3.4 – 3.6] предпринимались разнообразные попытки распространить идеи марковской фильтрации на двумерные сигналы. Прежде чем остановиться на одном из методов, развитых в работах [3.6, 3.8], рассмотрим кратко одномерную марковскую фильтрацию дискретных сигналов, поскольку она составляет основу двумерных процедур.

### 3.5.2. МАРКОВСКАЯ ФИЛЬТРАЦИЯ ОДНОМЕРНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Рассмотрим одномерную задачу фильтрации, когда входные данные представлены в виде одномерной последовательности наблюдений:

$$y_i = f(x_i, n_i), \quad i = \overline{0, I-1}. \quad (3.39)$$

Здесь все обозначения имеют тот же смысл, что и в (3.37) для двумерных сигналов. Для пояснения сути марковской фильтрации рассмотрим простейший вариант задачи: будем считать помеху независимым процессом (т.е.  $n_i$  — последовательность случайных независимых чисел), а сигнал — простой марковской последовательностью. На последнем остановимся подробнее. Последовательность является марковской, если ее совместное распределение вероятностей может быть представлено в виде

$$w_x(X) = w_x(x_0) \prod_{j=1}^{I-1} q(x_j | x_{j-1}). \quad (3.40)$$

Данное выражение содержит в правой части одномерное распределение  $w_x(x_0)$  для нулевого элемента последовательности и цепочку так называемых одношаговых распределений вероятностей перехода  $q(x_j | x_{j-1})$ , представляющих собой разновидность условных распределений. Соотношение (3.40) описывает свойство ограниченного последдействия, проявляющееся в том, что условное распределение  $q(x_j | x_{j-1})$  элемента  $x_j$  зависит лишь от единственного соседнего элемента  $x_{j-1}$ . Последовательность как бы «прошита» цепочкой непосредственных соседних связей. Элементы, удаленные друг от друга более чем на один шаг, непосредственным вероятностным механизмом не связаны. Это, впрочем, совсем не означает их независимости, зависимость проявляется опосредованно, через цепочку прямых связей.

Часто индексы  $i, j$ , входящие в (3.40), ассоциируют с дискретным временем, а последовательность  $x_i$  называют случайным процессом. Тогда о соотношении (3.40) говорят, что оно описывает процесс в прямом времени. Известно, что марковский процесс  $x_i$  обладает марковским свойством и в обратном времени, что позволяет записать его распределение вероятностей в виде

$$w_x(X) = w_x(x_{I-1}) \prod_{j=I-2}^0 \pi(x_j | x_{j+1}). \quad (3.41)$$

В соотношение (3.41) входят распределение последнего элемента и цепочка одношаговых распределений перехода в обратном времени  $\pi(x_j | x_{j+1})$ , не совпадающих с  $q(x_j | x_{j-1})$ .

Марковские процессы обладают разделяющим свойством, позволяющим представить их распределение еще в одной форме, полезной для разработки оптимальных процедур фильтрации. В соответствии с этим свойством любой элемент последовательности  $x_k$  разделяет ее на два множества  $X_{k-} = \{x_0, x_1, \dots, x_{k-1}\}$  и  $X_{k+} = \{x_{k+1}, x_{k+2}, \dots, x_{I-1}\}$ , которые при известном значении  $x_k$  условно независимы [3.7], т.е. имеют место равенства:

$$w(X_{k-} X_{k+} | x_k) = w(X_{k-} | x_k) w(X_{k+} | x_k),$$

$$\begin{aligned} w_x(X) &= w(x_k) w(X_{k-} | x_k) w(X_{k+} | x_k) = \\ &= w(x_k) \prod_{r=k-1}^0 \pi(x_r | x_{r+1}) \prod_{j=k+1}^{I-1} q(x_j | x_{j-1}). \end{aligned} \quad (3.42)$$

Последнее соотношение дает возможность построения некаузального фильтра, формирующего результат фильтрации при помощи очень удобных, экономичных вычислительных процедур. Это результат того, что апостериорное распределение вероятностей для произвольного  $k$ -го элемента последовательности можно представить в виде [3.7]

$$w(x_k | Y) = c_k \frac{w(x_k | Y_{k-}) w(x_k | Y_{k+})}{w(x_k | y_k)}. \quad (3.43)$$

В правую часть (3.43) входят три частных АРВ элемента  $x_k$ , различающихся составом входных данных, на которых основаны эти АРВ. Здесь  $Y_{k-} = \{y_0, y_1, \dots, y_{k-1}, y_k\}$ ,  $Y_{k+} = \{y_k, y_{k+1}, \dots, y_{I-1}\}$  — векторы прошлых и будущих данных соответственно, причем оба содержат текущий элемент  $y_k$ . Основанные на них в отдельности АРВ могли бы послужить базой для образования каузальной и антикаузальной оценок полезного сигнала. В знаменателе стоит одноточечное АРВ, компенсирующее двукратное при-

существование текущего наблюдения  $y_k$  в числителе, а коэффициент  $c_k$  подбирается так, чтобы обеспечивалась нормировка к единице получаемого АРВ.

Согласно (3.43) получение оценки состоит из двух этапов. На первом этапе из локальных входных данных формируются локальные АРВ, которые на втором этапе объединяют в окончательное АРВ, используемое далее для получения точечной оценки. Вычислительная сложность этого процесса в значительной степени определяется сложностью формирования локальных АРВ, главным образом находящихся в числителе формулы (3.43), так как получение одноточечного АРВ в знаменателе обычно – достаточно простая задача.

Определение локальных АРВ очень сильно облегчается при использовании марковских свойств последовательностей. Оказывается, что их можно вычислять при помощи рекуррентных соотношений. Так, например, АРВ  $w(x_k | Y_{k-})$  вычисляется на основе рекуррентного уравнения в прямом времени:

$$w(x_k | Y_{k-}) = c_{k-} \int_{\Omega_x} w(x_{k-1} | Y_{(k-1)-}) q(x_k | x_{k-1}) dx_{k-1}. \quad (3.44)$$

Здесь  $\Omega_x$  – область интегрирования, определяемая областью значений последовательности  $x_i$ , а  $c_{k-}$  – нормирующий коэффициент. Для вычисления текущего АРВ  $w(x_k | Y_{k-})$  необходимо, согласно (3.44), использовать АРВ  $w(x_{k-1} | Y_{(k-1)-})$ , являющееся одним из итогов работы фильтра в предыдущей точке. В удобстве этого выражения для практической реализации и состоит одно из основных преимуществ марковской фильтрации. Наряду с ограниченным объемом вычислений, предписанных (3.44), при работе на каждом шаге не требуется использовать полное множество входных данных, поскольку в выражение (3.44) входит только текущий элемент обработки  $y_k$ . Аналогично выглядит и рекуррентное соотношение для локального АРВ  $w(x_k | Y_{k+})$  с тем лишь отличием, что оно развивается в обратном времени от конца интервала наблюдения к его началу. Оба рекуррентных соотношения должны быть дополнены граничными условиями, определяющими одноточечные АРВ  $w(x_0 | y_0)$  и  $w(x_{I-1} | y_{I-1})$ , что, как упоминалось выше, не представляет сложной задачи.

В целом процедура фильтрации, основанная на приведенных соотношениях, выглядит следующим образом. Сначала выполняют обработку последовательности в прямом времени, в результате чего во всех точках формируется АРВ  $w(x_k | Y_{k-})$ , затем осуществляют развертку в обратном времени, в ходе которой формируются локальные АРВ  $w(x_k | Y_{k+})$ . Далее снова осуществляют развертку в прямом (или в обратном) времени, в ходе которой вычисляют одноточечные  $w(x_k | y_k)$  и полные  $w(x_k | Y)$  АРВ. Одновременно в каждой точке определяется точечная оценка, например, как координата максимума АРВ. Состав вычислений очень удобен для параллельной реализации. При наличии параллельных вычислительных устройств можно одновременно в различных вычислителях определять различные локальные АРВ и затем в режиме последовательной обработки объединять все локальные результаты. Само это объединение происходит независимо для всех точек последовательности, что позволяет финальную процедуру формирования окончательных точечных оценок выполнять снова параллельно во всех точках.

### 3.5.3. ДВУХЭТАПНАЯ МАРКОВСКАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ

Рассмотрим подход к фильтрации изображений, основу которого составляет использование двумерных, но неполных входных данных, а также наличие у них марковских свойств. Рассмотрим получение оценки изображения в произвольной точке кадра с координатами  $(i_1, i_2)$ , называя ее текущей точкой фильтрации. Осуществляя развертку изображения, можно последовательно в состояние текущей точки переводить все точки кадра, выполняя таким образом его полную обработку.

Будем считать, что для получения оценки  $x_{i_1, i_2}^*$  в точке  $(i_1, i_2)$  привлекаются лишь данные  $Y = \{Y^{(1)}, Y^{(2)}, Y^{(3)}, Y^{(4)}, y\}$ , образованные наблюдениями  $i_1$ -й строки и  $i_2$ -го столбца, на пересечении которых в кадре расположен оцениваемый элемент. Рис. 3.8 иллюстрирует геомет-

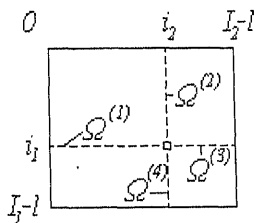


Рис. 3.8. Геометрия использования данных при двухэтапной фильтрации

рию задачи. Наблюдения  $Y^{(k)}$ ,  $k = \overline{1,4}$  образованы входными данными, лежащими на горизонтальных и вертикальных лучах  $\Omega^{(k)}$ , расходящихся из текущей точки. Причем в каждый из этих векторов не входит текущее наблюдение  $y = y_{i_1, i_2}$ , которое представлено в  $Y$  отдельным элементом.

Будем, кроме того, рассматривать такие случайные поля  $X$ , которые обладают свойством условной независимости. Это означает, что совместное распределение всех его элементов  $X$ , расположенных на «кресте»  $\Omega$  (рис. 3.8), можно представить в виде

$$w(X) = w(x) \cdot w(X^{(1)}, X^{(3)} | x) \cdot w(X^{(2)}, X^{(4)} | x), \quad (3.45)$$

где верхние индексы также указывают на принадлежность векторов соответствующим лучам. Соотношение (3.45) показывает, что значения сигнала на любой строке и на любом столбце изображения условно независимы, если известно значение сигнала  $x$  на пересечении этих строки и столбца. Если, кроме того, одномерные сигналы  $X^{(1)}, x, X^{(3)}$  и  $X^{(2)}, x, X^{(4)}$  являются марковскими последовательностями, для которых справедливо свойство условной независимости (3.42), то имеем:

$$w(X) = w(x) \cdot w(X^{(1)} | x) \cdot w(X^{(2)} | x) \cdot w(X^{(3)} | x) \cdot w(X^{(4)} | x). \quad (3.46)$$

Используя эту математическую модель изображения в случае независимой помехи  $n_{i_1, i_2}$ , можно одноточечное апостериорное распределение представить в виде [3.6]

$$w(x | Y) = c \cdot \frac{\prod_{k=1}^4 w(x | Y^{(k)}, y)}{w^3(x | y)}. \quad (3.47)$$

Соотношение (3.47) служит теоретической базой для построения оптимальных двухэтапных процедур фильтрации, использующих неполные данные исходных наблюдений. Полное АРВ, основанное на всех *привлекаемых при фильтрации* данных  $Y$ , как и в одномерном случае, представляется в виде произведения частных АРВ, каждое из которых использует локальные данные одного из лучей  $Y^{(k)}$  и текущий элемент  $y$ . Наличие в знаменателе третьей степени одноточечного АРВ служит компенсацией трехкратного «лишнего» участия текущего наблюдения в числителе. Константа  $c$  позволяет сделать АРВ нормированным.



Рассмотренный вариант обработки является разновидностью *некаузальной* фильтрации, поскольку в получении оценки участвуют элементы входного наблюдения, имеющие как большие, так и меньшие значения аргументов, чем у оцениваемого сигнала. Если в процессе обработки опираться только на данные двух лучей  $\Omega^{(1)}$  и  $\Omega^{(2)}$ , то будет получен двумерный *каузальный* фильтр. При добавлении к ним третьего луча  $\Omega^{(3)}$  реализуется полукаузальная фильтрация. Используя  $\Omega^{(1)}$  и  $\Omega^{(3)}$ , можем осуществить *одномерную некаузальную* фильтрацию. Коррекция, которой необходимо в этих случаях подвергнуть формулу (3.47), вполне очевидна.

Соотношение (3.47) дает возможность выполнить двумерную обработку изображения в виде некоторой совокупности одномерных процедур. Весь цикл вычислений можно представить следующим образом. Выполняется обработка всех строк изображения в прямом направлении (слева направо), в результате чего в каждой точке образуется распределение  $w(x|Y^{(1)}, y)$ . При этом используются одномерные рекуррентные процедуры, описанные выше. Далее происходит повторное сканирование строк, но в «обратном времени» — справа налево, в процессе которого вычисляются распределения  $w(x|Y^{(3)}, y)$ . Затем изображение аналогично дважды обрабатывается по столбцам — сверху вниз и снизу вверх, в результате чего определяются частные АРВ  $w(x|Y^{(2)}, y)$  и  $w(x|Y^{(4)}, y)$ . Вычислением одноточечного АРВ  $w(x|y)$  завершается первый этап обработки. На втором этапе происходит объединение всех частных АРВ в каждой точке кадра в окончательное АРВ, а также на его основе вычисляются точечные оценки изображений  $x_{i_1, i_2}^*$ .

С точки зрения скорости вычислений данная технология обработки очень привлекательна. Следует вместе с тем иметь в виду, что для ее реализации необходим достаточный запас оперативной памяти, чтобы хранить промежуточные результаты обработки, к числу которых относятся все частные АРВ. В этом отношении вычислительный процесс может быть существенно оптимизирован, поскольку ни одно из частных АРВ не представляет окончательной ценности. Это позволяет, например, не хранить отдельно пять различных распределений, входящих в правую часть (3.47), а по мере получения очередного сомножителя формировать произведение, именно которое и следует хранить в памяти до завершения вычислений. Очевидно, что структура вы-

числений, как и в одномерном случае, удобна для реализации при помощи многоканального вычислительного устройства.

Структура распределений очень сильно влияет на требуемые объем вычислений и ресурс памяти. Имеются очень «удобные» в этом смысле виды распределений. Например, если для описания изображения применима модель случайного поля с гауссовским распределением, то для представления каждого из частных и финального АРВ в (3.47) требуется наличие всего двух параметров — математического ожидания и дисперсии. Именно это и определяет конкретный характер и количество вычислений в процессе фильтрации, а также объем необходимой памяти.

Другим примером такого рода может служить математическая модель бинарного случайного поля, которое в различных точках принимает значения  $x_0$  или  $x_1$ . Такое описание также является очень экономичным, поскольку АРВ содержит всего две вероятности  $w(x_0|Y)$  и  $w(x_1|Y)$ , непосредственное вычисление которых и выполняется при помощи (3.47).

Существует отдельный вопрос, связанный с применимостью марковских двумерных моделей (3.45), (3.46), позволяющих построить эффективные двухэтапные процедуры. Его изучение является достаточно непростой теоретической задачей. В частности, в работах [3.6, 3.8] установлено, что и для гауссовских, и для бинарных случайных полей необходимым и достаточным условием применимости (3.45) является возможность представления двумерных корреляционных функций этих полей в *разделимом виде*, т.е. в виде произведения двух множителей, один из которых описывает корреляцию изображения по строке, а второй — по столбцу. Дополнительные требования, вытекающие из (3.46), сводятся к существованию марковских свойств у одномерных последовательностей в горизонтальном и вертикальном сечении изображения. В двух указанных примерах наличие таких свойств связано с *экспоненциальным видом* корреляционных функций этих одномерных сечений изображения.

На рис. 3.9 приведены результаты экспериментальной проверки двумерных двухэтапных алгоритмов фильтрации изображения. На рис. 3.9, *а* показано тестовое бинарное изображение «острова», на рис. 3.9, *б* — изображение, искаженное белым гауссовским шумом (отношение сигнал/шум  $q^2 = 0$  дБ). Рис. 3.9, *в* иллюстрирует применение простой поэлементной пороговой обработки (см. рис. 1.4, *а*), при которой порог определялся так, чтобы реализовывалась одноточечная процедура максимума апостериорной вероятности. На рис. 3.9, *г*, 3.9, *д* и 3.9, *е* показаны

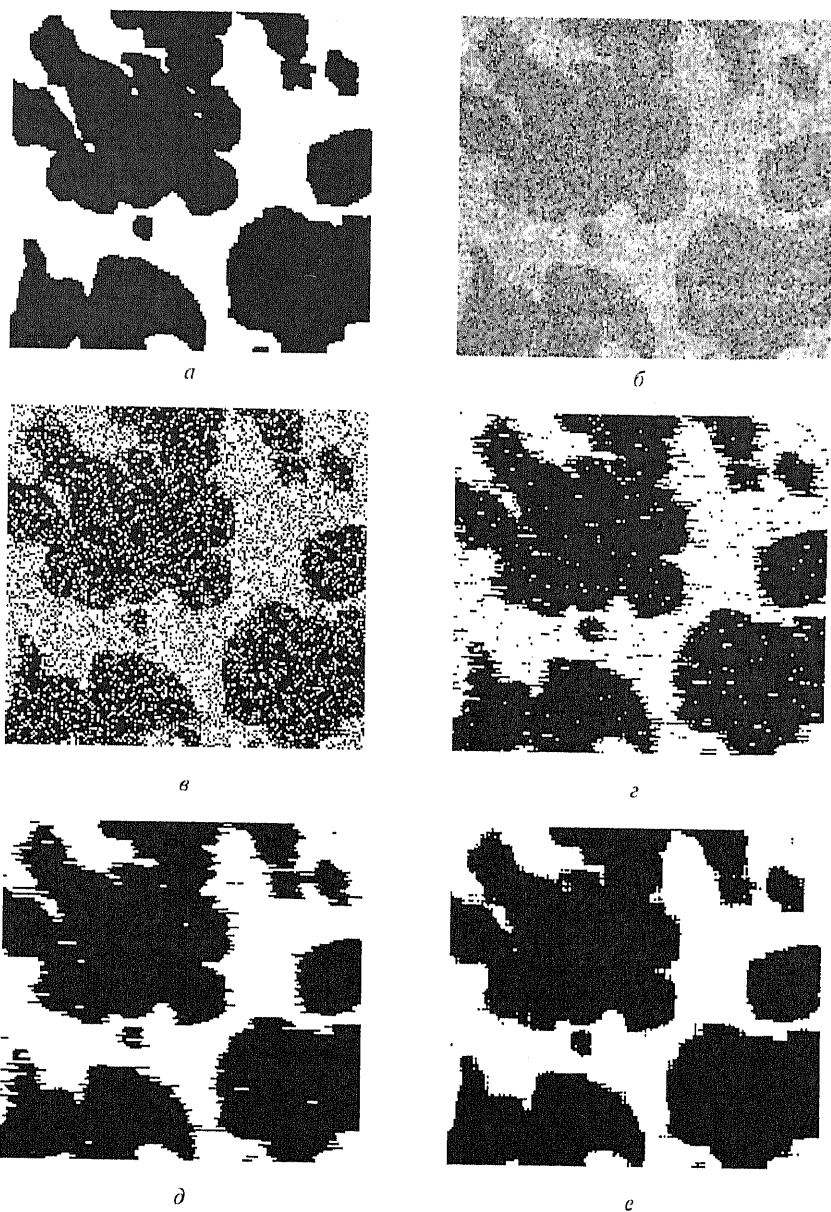


Рис. 3.9. Двухэтапная марковская фильтрация изображения

различные результаты двухэтапной фильтрации. Первый из них соответствует одномерной каузальной фильтрации, второй — также одномерной, но некаузальной, а третий — двумерной некаузальной процедуре. Визуальное сравнение результатов говорит об очень низком качестве поэлементной обработки. При ее использовании вероятность ошибки (т.е. события, состоящего в замене числа  $x_0$  числом  $x_1$  или наоборот) составила 0.23. Качество обработки улучшается при использовании фильтрации, причем оно повышается как при переходе от одномерной каузальной (при которой вероятность ошибки составляет 0.086) к одномерной некаузальной (вероятность ошибки 0.041), так и при переходе к двумерной обработке, при которой достигается вероятность ошибки, равная 0.022. Таким образом, применение одномерной некаузальной фильтрации позволяет уменьшить вероятность ошибки в 5 раз по сравнению с поэлементной пороговой обработкой, а двумерной некаузальной фильтрации — почти в 10 раз. Эти примеры говорят об очень высокой эффективности, которой может достигать фильтрация, и убеждают в полезности тех значительных усилий, которые необходимы для нахождения эффективных алгоритмов.

### 3.6. МЕДИАННАЯ ФИЛЬТРАЦИЯ

Все линейные алгоритмы фильтрации приводят к сглаживанию резких перепадов яркости изображений, прошедших обработку. Этот недостаток, особенно существенный, если потребителем информации является человек, принципиально не может быть устранен в рамках линейной обработки. Дело в том, что линейные процедуры являются оптимальными при гауссовском распределении сигналов, помех и наблюдаемых данных. Реальные изображения, строго говоря, не подчиняются данному распределению вероятностей. Причем одна из основных причин этого состоит в наличии у изображений разнообразных границ, перепадов яркости, переходов от одной текстуры к другой и т.п. Поддаваясь локальному гауссовскому описанию в пределах ограниченных участков, многие реальные изображения в этой связи плохо представляются как глобально гауссовские объекты. Именно это и служит причиной плохой передачи границ при линейной фильтрации.

Вторая особенность линейной фильтрации — ее оптимальность, как только что упоминалось, при гауссовском характере помех. Обычно этому условию отвечают шумовые помехи на

изображениях, поэтому при их подавлении линейные алгоритмы имеют высокие показатели. Однако часто приходится иметь дело с изображениями, искаженными помехами других типов. Одной из них является импульсная помеха. При ее воздействии на изображении наблюдаются белые или (и) черные точки, хаотически разбросанные по кадру. Применение линейной фильтрации в этом случае неэффективно — каждый из входных импульсов (по сути дельта-функция) дает отклик в виде импульсной характеристики фильтра, а их совокупность способствует распространению помехи на всю площадь кадра.

Удачное решение перечисленных проблем — применение *медианной фильтрации*, предложенной Дж. Тьюки в 1971 г. для анализа экономических процессов. Наиболее полное исследование медианной фильтрации применительно к обработке изображений приведено в сборнике [3.9]. Отметим, что медианная фильтрация представляет собой эвристический метод обработки, ее алгоритм не является математическим решением строго сформулированной задачи. Поэтому исследователи уделяют большое внимание *анализу эффективности* обработки изображений на ее основе и сопоставлению с другими методами.

При применении медианного фильтра (МФ) происходит последовательная обработка каждой точки кадра, в результате чего образуется последовательность оценок. В идейном отношении обработка в различных точках независима (этим МФ похож на масочный фильтр), но в целях ее ускорения целесообразно алгоритмически на каждом шаге использовать ранее выполненные вычисления.

При медианной фильтрации используется двумерное окно (апертура фильтра), обычно имеющее центральную симметрию, при этом его центр располагается в текущей точке фильтрации. На рис. 3.10 показаны два примера наиболее часто применяемых вариантов окон в виде креста и в виде квадрата. Размеры апер-

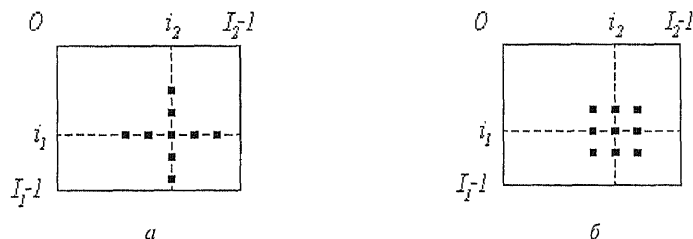


Рис. 3.10. Примеры окон при медианной фильтрации

туры принадлежат к числу параметров, оптимизируемых в процессе анализа эффективности алгоритма. Отсчеты изображения, оказавшиеся в пределах окна, образуют *рабочую выборку* текущего шага.

Двумерный характер окна позволяет выполнять, по существу, двумерную фильтрацию, поскольку для образования оценки привлекаются данные как из текущих строки и столбца, так и из соседних. Обозначим рабочую выборку в виде одномерного массива  $Y = \{y_1, y_2, \dots, y_n\}$ ; число его элементов равняется размеру окна, а их расположение произвольно. Обычно применяют окна с нечетным числом точек  $n$  (это автоматически обеспечивается при центральной симметрии апертуры и при вхождении самой центральной точки в ее состав). Если упорядочить последовательность  $\{y_i, i = \overline{1, n}\}$  по возрастанию, то ее *медианой* будет тот элемент выборки, который занимает центральное положение в этой упорядоченной последовательности. Полученное таким образом число и является продуктом фильтрации для текущей точки кадра. Понятно, что результат такой обработки в самом деле не зависит от того, в какой последовательности представлены элементы изображения в рабочей выборке  $Y$ . Введем формальное обозначение описанной процедуры в виде

$$x^* = \text{med}(y_1, y_2, \dots, y_n). \quad (3.48)$$

Рассмотрим пример. Предположим, что выборка имеет вид  $Y = \{136, 110, 99, 45, 250, 55, 158, 104, 75\}$ , а элемент 250, расположенный в ее центре, соответствует текущей точке фильтрации  $(i_1, i_2)$  (рис. 3.10). Большое значение яркости в этой точке кадра может быть результатом воздействия импульсной (точечной) помехи. Упорядоченная по возрастанию выборка имеет при этом вид  $\{45, 55, 75, 99, 104, 110, 136, 158, 250\}$ , следовательно, в соответствии с процедурой (3.48), получаем  $x^* = \text{med}(y_1, y_2, \dots, y_9) = 104$ . Видим, что влияние «соседей» на результат фильтрации в текущей точке привело к «игнорированию» импульсного выброса яркости, что следует рассматривать как эффект фильтрации. Если импульсная помеха не является точечной, а покрывает некоторую локальную область, то она также может быть подавлена. Это произойдет, если размер локальной области будет меньше, чем половина размера апертуры МФ. Поэтому для подавления импульсных помех, поражающих локальные участки изображения, надо увеличивать размеры апертуры МФ.

Из (3.48) следует, что действие МФ состоит в «игнорировании» экстремальных значений входной выборки — как положительных, так и отрицательных выбросов. Такой принцип подавления помехи может быть применен и для ослабления шума на изображении. Однако исследование подавления шума при помощи медианной фильтрации показывает, что ее эффективность при решении этой задачи ниже, чем у линейной фильтрации [3.9].

Результаты экспериментов, иллюстрирующие работу МФ, приведены на рис. 3.11. В экспериментах применялся МФ, имеющий квадратную апертуру со стороной, равной 3. В левом ряду представлены изображения, искаженные помехой, в правом — результаты их медианной фильтрации. На рис. 3.11, *а* и 3.11, *в* показано исходное изображение, искаженное импульсной помехой. При ее наложении использовался датчик случайных чисел с равномерным на интервале  $[0,1]$  законом распределения, вырабатывающий во всех точках кадра независимые случайные числа. Интенсивность помехи задавалась вероятностью  $p$  ее возникновения в каждой точке. Если для случайного числа  $n_{i_1, i_2}$ , сформированного в точке  $(i_1, i_2)$ , выполнялось условие  $n_{i_1, i_2} < p$ , то яркость изображения  $x_{i_1, i_2}$  в этой точке замещалась числом 255, соответствующим максимальной яркости (уровню белого). На рис. 3.11, *а* действием импульсной помехи искажено 5 % ( $p = 0.05$ ), а на рис. 3.11, *в* — 10 % элементов изображения. Результаты обработки говорят о практически полном подавлении помехи в первом случае и о ее значительном ослаблении во втором.

Рис. 3.11, *д* показывает изображение, искаженное независимым гауссовским шумом при отношении сигнал/шум  $q^2 = -5$  дБ, а рис. 3.11, *е* — результат его фильтрации медианным фильтром. Условия данного эксперимента позволяют сравнивать его результаты с результатами рассмотренной выше линейной фильтрации. В табл. 3.1 приведены показатели, дающие возможность такого сравнения. Для различных методов фильтрации в этой таблице приводятся значения относительного среднего квадрата ошибок  $\delta_\epsilon^2$  и коэффициента ослабления шума  $\gamma$  для случая, когда отношение сигнал/шум на входе фильтра составляет  $-5$  дБ.

Наибольшей эффективностью обладает двумерный фильтр Винера, уменьшающий средний квадрат ошибок в 17 раз. Медианный фильтр имеет наименьшую из всех рассмотренных фильтров эффективность, ему соответствует  $\gamma = 5.86$ . Тем не менее это число свидетельствует о том, что и при его помощи удастся значительно снизить уровень шума на изображении.

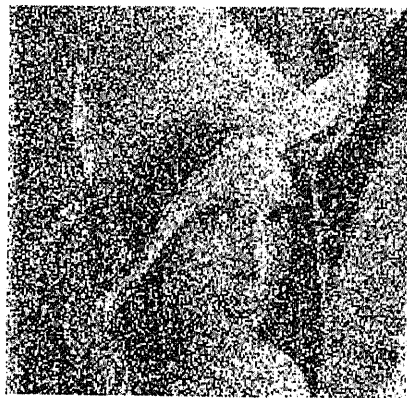
*a**б**в**г**д**е*

Рис. 3.11. Примеры медианной фильтрации



Т а б л и ц а 3.1

**Сравнение эффективности подавления шума  
при фильтрации изображений,  $q^2 = -5$  дБ**

| Показатели            | Масочный фильтр с оптимальной КИХ | Масочный фильтр с равномерной КИХ | Двумерный рекуррентный фильтр | Двумерный фильтр Винера | Медианный фильтр |
|-----------------------|-----------------------------------|-----------------------------------|-------------------------------|-------------------------|------------------|
| $\delta_{\epsilon}^2$ | 0.309                             | 0.395                             | 0.29                          | 0.186                   | 0.539            |
| $\gamma$              | 10.2                              | 8.0                               | 10.9                          | 17.0                    | 5.86             |

Вместе с тем, как говорилось выше, и что демонстрирует рис. 3.11,  $\epsilon$ , медианная фильтрация в меньшей степени сглаживает границы изображения, чем любая линейная фильтрация. Механизм этого явления очень прост и заключается в следующем. Предположим, что апертура фильтра находится вблизи границы, разделяющей светлый и темный участки изображения, при этом ее центр располагается в области темного участка. Тогда, вероятнее всего, рабочая выборка будет содержать большее количество элементов с малыми значениями яркости и, следовательно, медиана будет находиться среди тех элементов рабочей выборки, которые соответствуют этой области изображения. Ситуация меняется на противоположную, если центр апертуры смещен в область более высокой яркости.

### ВОПРОСЫ К ГЛАВЕ 3

3.1. Докажите, что при наблюдении аддитивной смеси независимых изображения и шума взаимная корреляционная функция  $B_{xy}(i, j)$  совпадает с корреляционной функцией изображения  $B_x(i, j)$ .

3.2. Приведите уравнение Винера-Хопфа (3.6) к виду (3.11) применительно к случаю аддитивного взаимодействия изображения и белого шума.

3.3. Докажите справедливость выражения (3.7) для среднего квадрата ошибок фильтрации.

3.4. Поясните, почему при ограниченном размере окрестности, применяемой при КИХ-фильтрации, нельзя достичь предельного подавления шума.

3.5. Докажите, что стационарный фильтр Калмана, работающий в соответствии с алгоритмом (3.12), имеет импульсную характеристику (3.13).

3.6. Докажите, что при оптимальной линейной фильтрации ошибка и результат фильтрации ортогональны.

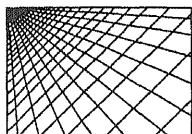
3.7. Получите алгебраическое уравнение (3.18) для коэффициента усиления двумерного рекуррентного фильтра и качественно исследуйте тенденцию изменения его решения при вариации отношения сигнал/шум.

3.8. Приведите алгоритм фильтрации в двумерном фильтре с ИХ (3.14) к рекуррентному виду (3.19).

3.9. Докажите тождественность прямого и обратного двумерных ДПФ.

3.10. Покажите на примере, как может быть получено одноточечное апостериорное распределение вероятностей, входящее в формулу (3.43), для АРВ марковской последовательности.

3.11. По аналогии с рекуррентным уравнением прямого времени для локального АРВ  $w(x_k(Y_{k-}))$  запишите рекуррентное уравнение обратного времени для АРВ  $w(x_k(Y_{k+}))$  и поясните его.



## ГЛАВА 4

### ВОССТАНОВЛЕНИЕ ИЗОБРАЖЕНИЙ

Из-за несовершенства формирующих и регистрирующих систем записанное ими изображение представляет собой искаженную (нечеткую) копию оригинала. Основными причинами искажений, приводящих к ухудшению четкости, являются ограниченная разрешающая способность формирующей системы, расфокусировка, наличие искажающей среды (например, атмосферы), движение камеры по отношению к регистрируемому объекту и т.п. Устранение или ослабление искажений с целью повышения резкости относится к задаче восстановления изображений.

Наиболее общая схема формирования изображения представлена на рис. 4.1, где  $u(v, v)$  — неизвестная функция распределения яркости объекта, описываемая функцией двух переменных  $(v, v)$ ;  $s(x, y)$  — наблюдаемое изображение, сформированное из

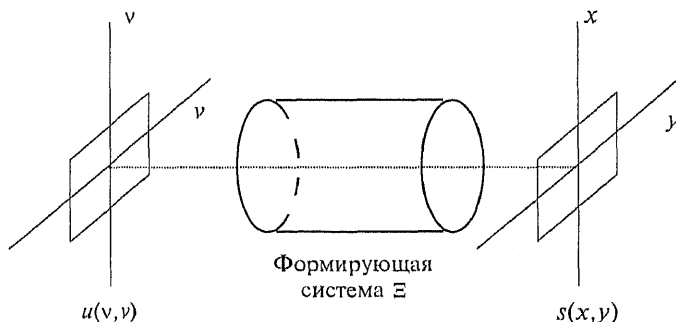


Рис. 4.1. Схема формирования изображения

$u(v, v)$  при помощи некоторого известного оператора искажений  $\Xi$ :

$$s(x, y) = \Xi u(v, v).$$

Вид оператора  $\Xi$  определяется свойствами формирующей системы. Двумерную функцию  $u(v, v)$  в дальнейшем будем называть исходным изображением. Задача восстановления заключается в нахождении изображения  $\hat{u}(v, v)$ , являющегося оценкой исходного изображения  $u(v, v)$  по наблюдаемому изображению  $s(x, y)$ , т.е. в устранении искажений, вносимых оператором  $\Xi$ .

#### 4.1. МОДЕЛИ ИЗОБРАЖЕНИЙ И ИХ ЛИНЕЙНЫХ ИСКАЖЕНИЙ

##### 4.1.1. ФОРМИРОВАНИЕ ИЗОБРАЖЕНИЙ

Большинство формирующих систем в первом приближении можно рассматривать как линейные и инвариантные к сдвигу. Изображения, сформированные такими системами, претерпевают *линейные пространственно-инвариантные искажения*, характеризующиеся тем, что механизм их возникновения один и тот же для всех точек  $(x, y)$ . Линейные искажения проявляются в ослаблении верхних частот исходного изображения. Визуально это приводит к ухудшению его резкости. В процессе записи изображения искажаются также шумами, присутствующими в любом реальном физическом устройстве. В ряде практически важных случаев шум можно считать аддитивным и независимым от исходного изображения.

С учетом изложенного выше наблюдаемое нерезкое изображение  $s(x, y)$  можно представить как выход линейной системы, показанной на рис. 4.2, а математическая модель процесса его формирования имеет вид

$$s(x, y) = z(x, y) + n(x, y),$$

где  $n(x, y)$  — аддитивный двумерный шум. Изображение  $z(x, y)$ , являющееся линейным искажением исходного изображения при отсутствии шума, определяется интегралом свертки:

$$\begin{aligned} z(x, y) &= h(x, y) \otimes u(x, y) = \iint_{(v, v) \in \Theta^{(u)}} h(x - v, y - v) u(v, v) dv dv = \\ &= \iint_{(v, v) \in \Theta^{(h)}} h(v, v) u(x - v, y - v) dv dv, \quad (x, y) \in \Theta^{(s)}, \end{aligned} \quad (4.1)$$

где  $\otimes \otimes$  — символ двумерной свертки;  $h(x, y)$  — двумерная импульсная характеристика (или ФРТ — функция рассеяния точки) линейной искажающей системы. Таким образом, значение функции яркости  $u(v, v)$  исходного изображения в точке с координатами  $(v, v)$  «размазывается» в соответствии с видом ФРТ  $h(x, y)$  и искажается аддитивным шумом.

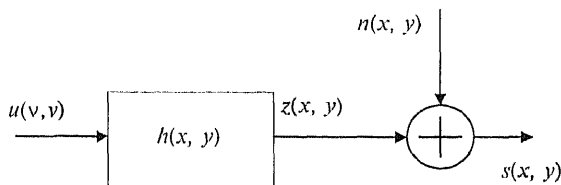


Рис. 4.2. Линейная модель формирования изображения

В выражении (4.1) учтено, что изображения, встречающиеся в практических задачах, имеют конечные размеры. Это означает, что яркость изображения полагается равной нулю всюду, кроме некоторой конечной области, которую будем называть кадром и обозначать через  $\Theta^{(\cdot)}$ . Верхний индекс в кадре  $\Theta^{(\cdot)}$  соответствует символу изображения или ФРТ, для которого определен этот кадр. Например, кадр изображения  $s(x, y)$  будем обозначать через  $\Theta^{(s)}$ . Мы будем рассматривать только прямоугольные кадры, стороны которых параллельны координатным осям. Под размерами кадра будем понимать совокупность длин вертикальной  $L_x^{(s)}$  и горизонтальной  $L_y^{(s)}$  сторон кадра  $\Theta^{(s)}$ . Часто оказывается удобным совместить центр кадра изображения с началом координат на плоскости изображения, тогда

$$s(x, y) = 0 \quad \text{при} \quad |x| \geq L_x^{(s)} / 2, \quad |y| \geq L_y^{(s)} / 2, \quad (4.2)$$

или когда координаты  $(x, y) \notin \Theta^{(s)}$ .

Относительные размеры кадров изображений и ФРТ в модели формирования (4.1) имеют важное значение. Как будет показано ниже, конечность их размеров значительно усложняет решение задачи восстановления. Размеры кадров  $\Theta^{(s)}$ ,  $\Theta^{(z)}$  и  $\Theta^{(n)}$  равны между собой и обусловлены тем, что любая формирующая система имеет ограниченное поле зрения (рис. 4.3).

Кадр  $\Theta^{(h)}$  функции рассеяния точки представляет собой минимальный прямоугольник, содержащий все точки с координатами  $(x, y)$ , в которых величина абсолютного значения ФРТ

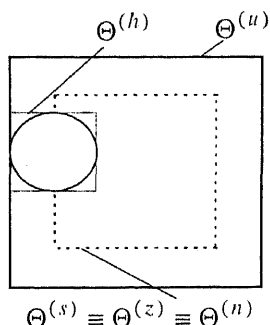


Рис. 4.3. Относительные размеры изображения и ФРТ

заметно отлична от нуля, например, когда  $|h(x, y)| \geq 10^{-10} \max(|h(x, y)|)$ . Следует отметить, что кадр  $\Theta^{(h)}$  ФРТ не всегда симметричен относительно начала координат. В частности, симметрией ФРТ не обладают каузальные линейные системы. Действие ФРТ сводится к тому, что каждая точка исходного изображения  $u(v, v)$  «размывается» в некоторую область, ограниченную кадром  $\Theta^{(h)}$ . Кадр  $\Theta^{(u)}$  исходного изображения может быть построен путем перемещения  $\Theta^{(h)}$  и представляет собой область всех точек, охватываемых кадром  $\Theta^{(h)}$  при его перемещении по кадру  $\Theta^{(s)}$  наблюдаемого

изображения. Поэтому, даже если регистрируемый объект имеет бесконечные размеры, наблюдаемое изображение формируется лишь за счет некоторой его части. Причем размеры кадра наблюдаемого изображения всегда меньше размеров исходного или равны им. Размеры кадров исходного и наблюдаемого изображений равны лишь при отсутствии линейных искажений, т.е. когда импульсная характеристика искажающей системы равна дельта-функции. Нас будет интересовать восстановление изображения в пределах кадра  $\Theta^{(s)}$ . В некоторых случаях удастся восстановить изображение в пределах кадра  $\Theta^{(u)}$  исходного изображения, так как та его часть, которая лежит вне пределов кадра  $\Theta^{(s)}$ , также оказывает влияние на наблюдаемое изображение  $s(x, y)$ .

Для изображений, представленных в цифровой форме, двумерные функции  $u(v, v)$ ,  $z(x, y)$ ,  $s(x, y)$ ,  $h(x, y)$  и  $n(x, y)$  с непрерывными аргументами в (4.1) заменяются двумерными массивами отсчетов, взятых на прямоугольных решетках с одинаковыми расстояниями  $\Delta = \Delta x = \Delta y$  между узлами. В этом случае модель наблюдаемого изображения имеет вид

$$s(i_1, i_2) = z(i_1, i_2) + n(i_1, i_2), \quad i_1 = -\frac{L_1^{(s)}}{2}, \frac{L_1^{(s)}}{2}, \quad i_2 = -\frac{L_2^{(s)}}{2}, \frac{L_2^{(s)}}{2}, \quad (4.3)$$

где

$$z(i_1, i_2) = h(i_1, i_2) \otimes u(i_1, i_2) = \sum_{(j_1, j_2) \in \Theta^{(h)}} h(j_1, j_2) u(i_1 - j_1, i_2 - j_2). \quad (4.4)$$

Аргументы с индексом 1 обозначают номер строки, а с индексом 2 — номер столбца. В дискретном случае размеры кадра (число отсчетов)  $L_{i_1}^{(v)}$  и  $L_{i_2}^{(v)}$  определяются отношением длин соответственно вертикальной и горизонтальной сторон кадра аналогового изображения к величине интервала дискретизации  $\Delta$ .

Операция свертки, которая имеется в формулах (4.1) и (4.4), эквивалентна умножению в частотной области. Это позволяет выполнить быструю имитацию линейных искажений с помощью ДПФ, заменив обычную свертку циклической (смотри главу 3). Как правило, размеры кадра ФРТ много меньше размеров кадра исходного изображения, поэтому перед преобразованием массив  $h(i_1, i_2)$  должен быть дополнен нулями. Кроме того, полагается, что изображения и ФРТ являются периодически продолженными, которые, так же как и в главе 3, обозначим волнистой линией.

Спектр линейно искаженного изображения  $\tilde{z}(i_1, i_2)$  равен произведению спектра  $\tilde{U}(k_1, k_2)$  исходного изображения  $\tilde{u}(i_1, i_2)$  и передаточной функции  $\tilde{H}(k_1, k_2)$  искажающей системы:

$$\tilde{Z}(k_1, k_2) = \tilde{U}(k_1, k_2) \tilde{H}(k_1, k_2), \quad (4.5)$$

где  $k_1$  и  $k_2$  — пространственные частоты. Размеры кадра изображения  $\tilde{z}(i_1, i_2)$ , полученного после обратного ДПФ от  $\tilde{Z}(k_1, k_2)$ , равны размерам кадра  $\Theta^{(u)}$  исходного изображения. Для завершения процедуры имитации необходимо «обрезать» края изображения  $\tilde{z}(i_1, i_2)$  до размеров кадра  $\Theta^{(s)}$  и добавить аддитивную помеху  $n(i_1, i_2)$ . Здесь и далее результаты преобразования Фурье от функций в пространственной области будут обозначаться соответствующими прописными буквами.

Рассмотрим импульсные и частотные характеристики формирующих систем при смазе и расфокусировке.

#### 4.1.2. РАЗМЫТИЕ ВСЛЕДСТВИЕ ДВИЖЕНИЯ (СМАЗ)

Смаз изображения возникает при взаимном движении камеры и объекта относительно друг друга во время экспозиции. Наблюдаемое изображение окажется как бы результатом наложения

со смещением множества исходных изображений. Мы рассмотрим только тот случай, когда камера перемещается с постоянной горизонтальной скоростью относительно снимаемого объекта. ФРТ и передаточная функция такой системы определяются выражениями:

$$h(x, y) = \begin{cases} \frac{1}{\alpha} \delta(x), & \text{при } 0 \leq y \leq \alpha; \\ 0 & \text{в остальных случаях,} \end{cases}$$

$$H(\omega_x, \omega_y) = \frac{\sin(\alpha \omega_y)}{\omega_y},$$

где длина смаза  $\alpha$  равна произведению скорости движения камеры на время экспозиции. Соответственно в дискретном случае ФРТ смаза будет

$$h(i_1, i_2) = \begin{cases} \frac{1}{L_{i_2}^{(h)}}, & \text{при } 0 \leq i_2 \leq L_{i_2}^{(h)} - 1 \text{ и } i_1 = 0; \\ 0 & \text{в остальных случаях,} \end{cases} \quad (4.6)$$

где размеры кадра  $L_{i_2}^{(h)} = [\alpha / \Delta]$  и  $L_{i_1}^{(h)} = 1$ . Здесь квадратные скобки обозначают операцию округления до целого. Взаимное расположение кадров изображений и ФРТ при смазе показано на рис. 4.4.

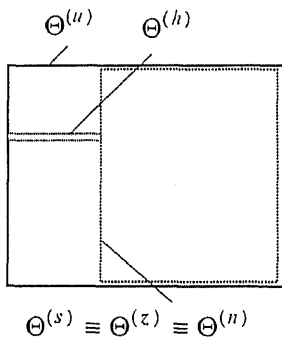


Рис. 4.4. Взаимное расположение изображения и ФРТ при смазе

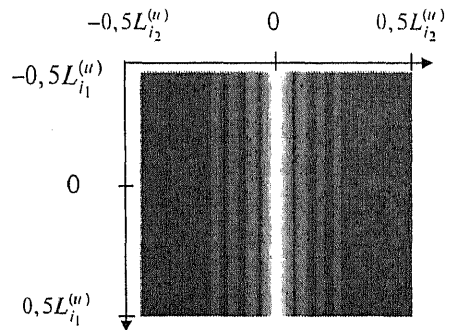


Рис. 4.5. Изображение модуля частотной характеристики искажающей системы



Дополняя ФРТ (4.6) нулями до размеров кадра исходного изображения и применяя двумерное ДПФ, получим частотную характеристику искажающей системы:

$$\tilde{H}(k_1, k_2) = \frac{1}{L_{i_2}^{(h)} \sqrt{L_{i_2}^{(u)}}} e^{j(\pi(L_{i_2}^{(h)} - 1)k_2 / L_{i_2}^{(u)})} \frac{\sin(\pi L_{i_2}^{(h)} k_2 / L_{i_2}^{(u)})}{\sin(\pi k_2 / L_{i_2}^{(u)})} \quad (4.7)$$

для всех  $k_1$ .

Изображение модуля  $\tilde{H}(k_1, k_2)$  показано на рис. 4.5 при  $L_{i_2}^{(h)} = 15$  и размерах исходного изображения  $185 \times 170$  элементов.

На рис. 4.7 приведен искаженный вариант исходного изображения «Сатурн» (рис. 4.6). Горизонтальный смаз составляет 15 элементов. Исходное изображение содержит  $185 \times 170$  элементов, а искаженное  $170 \times 170$  элементов.

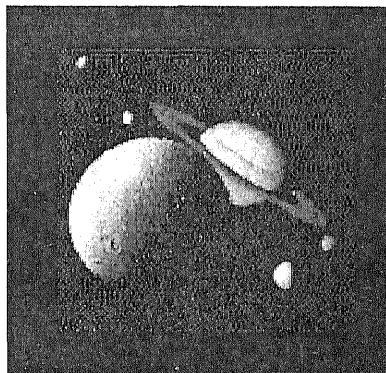


Рис. 4.6. Исходное изображение  
«Сатурн»

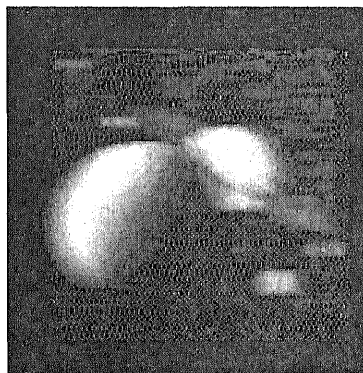


Рис. 4.7. Смазанное изображение  
«Сатурн»

#### 4.1.3. РАСФОКУСИРОВКА

Четкость изображения характеризуется воспроизведением мелких деталей и определяется разрешающей способностью формирующей системы. Разрешающая способность, например оптической системы, численно выражается количеством пар черно-белых линий на 1 мм изображения, которое формируется объективом системы. Если плоскость формируемого изображения находится в фокусе объектива, то пучок лучей, исходящий от точки на объекте, сходится в точку на изображении. При расфокусировке точка воспроизводится в виде некоторого пятна

(кружка размытия) и две близко расположенные точки на исходном изображении сливаются в одну на наблюдаемом. Величина кружка размытия зависит от фокусного расстояния объектива, а также от расстояний от объектива до объекта и до плоскости формируемого изображения [4.1]. Дискретное изображение будет четким (сфокусированным), если диаметр кружка размытия не превышает шага дискретизации  $\Delta$  наблюдаемого изображения. В противном случае линейные искажения становятся заметными.

При расфокусировке распределение интенсивности на изображении точечного источника, формируемого тонкой линзой с круговой апертурой, постоянно в пределах кружка размытия радиусом  $r$  и равно нулю за его пределами. Это соответствует цилиндрической ФРТ:

$$h(x, y) = \begin{cases} \frac{1}{\pi r^2}, & x^2 + y^2 \leq r^2; \\ 0 & \text{в остальных случаях.} \end{cases} \quad (4.8)$$

Из (4.8) следует, что размеры кадра  $L_x^{(h)} = L_y^{(h)} = 2r$ . Взяв двумерное преобразование Фурье от (4.8), получим передаточную функцию оптической системы

$$H(\omega_x, \omega_y) = J_1(r\rho) / r\rho, \quad \rho^2 = \omega_x^2 + \omega_y^2, \quad (4.9)$$

где  $J_1(\cdot)$  — функция Бесселя первого порядка. В дискретном случае ФРТ (4.8) имеет вид

$$h(i_1, i_2) = \begin{cases} \frac{1}{\pi r^2}, & i_1^2 + i_2^2 \leq r^2; \\ 0 & \text{в остальных случаях.} \end{cases} \quad (4.10)$$

На рис. 4.8 и 4.9 показаны ФРТ для тонкой линзы (4.10) и модуль ее передаточной функции при радиусе кружка размытия  $r = 10$  и размерах кадра изображения  $200 \times 200$  элементов.

Соотношения между различными кадрами изображений при расфокусировке соответствуют тем, что приведены на рис. 4.3, поскольку кадр ФРТ симметричен относительно центра координат. На рис. 4.12 показан дефокусированный вариант изображения «Сатурн» (см. рис. 4.6). Свертка исходного изображения производилась с гауссовским импульсом при  $\alpha = 0.1$ . Искаженное изображение содержит  $170 \times 170$  элементов.

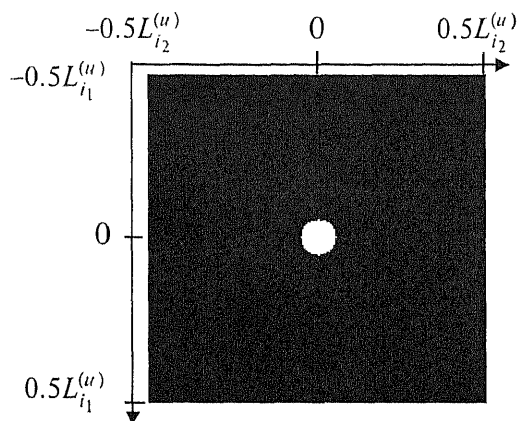


Рис. 4.8. ФРТ тонкой линзы

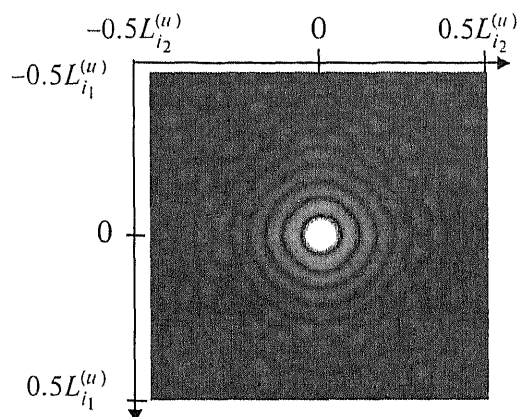


Рис. 4.9. Изображение модуля частотной характеристики тонкой линзы

Земную атмосферу также можно рассматривать как оптическую систему. В качестве приближенной модели ФРТ такой системы используется двумерный гауссовский импульс

$$h(x, y) = K \exp[-\alpha (x^2 + y^2)], \quad (4.11)$$

который в дискретном случае имеет вид

$$h(i_1, i_2) = K \exp[-\alpha (i_1^2 + i_2^2)], \quad (4.12)$$

где  $K$  — нормирующий коэффициент,  $\alpha$  — коэффициент пространственной нерезкости. Передаточная функция, соответствующая ФРТ (4.11), определяется выражением

$$H(\omega_x, \omega_y) = \exp\left[-(\omega_x^2 + \omega_y^2) / 4\alpha\right]. \quad (4.13)$$

Очевидно, что точки, для которых выполняется условие (4.2), образуют круг радиусом

$$r = \sqrt{-\frac{\ln(10^{-10})}{\alpha}}. \quad (4.14)$$

Следовательно, чем больше  $\alpha$ , тем меньше расфокусировка наблюдаемого изображения. ФРТ для земной атмосферы и соответствующая ей передаточная функция при  $\alpha = 0.1$  показаны на рис. 4.10 и 4.11. Радиус кружка размытия примерно равен 15. Размеры пятна ФРТ на рис. 4.10 визуально кажутся меньше, чем размеры пятна для тонкой линзы (см. рис. 4.8), так как гауссовский импульс является быстро убывающей функцией.

Таким образом, можно выделить три основных фактора, которые существенно усложняют решение проблемы восстановления изображений.

1. Искажения типа «расфокусировка» или «смаз» проявляются в ослаблении верхних пространственных частот изображения, так как формирующие системы представляют собой фильтры нижних частот. При этом отношение сигнал/шум на верхних частотах, определяющих четкость изображения, будет значительно хуже, чем для изображения в целом. Если система, формирующая изображение, ослабляет сигнал на каких-то пространственных частотах, то при восстановлении он должен быть усилен в той мере, в какой был ослаблен. Вместе с сигналом будут усиливаться и шумы. Поэтому улучшение качества изображения по резкости может привести к ухудшению его качества по зашумленности.

2. Яркость на краях кадра искаженного изображения зависит от яркости объектов, расположенных вне кадра, за счет свертки исходного изображения с ФРТ. При восстановлении изображений из-за неполной информации о сигнале вне кадра возникают краевые эффекты. Влияние краевых эффектов на качество восстановления в ряде случаев оказывается даже более существенным, чем зашумленность изображения.

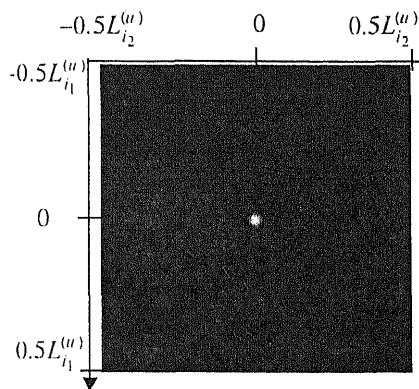


Рис. 4.10. ФРТ атмосферы Земли

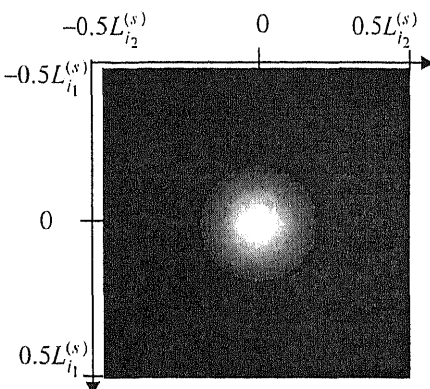


Рис. 4.11. Изображение модуля частотной характеристики атмосферы Земли

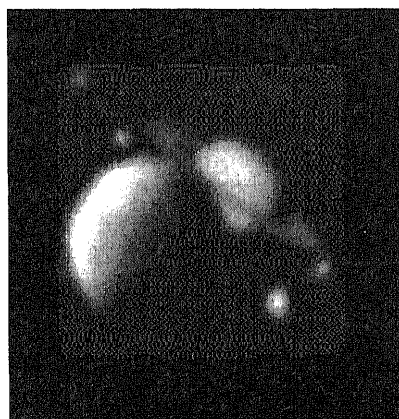


Рис. 4.12. Дефокусированное изображение «Сатурн»

3. При искажениях, вызванных движением или расфокусировкой камеры, передаточные функции (4.7) и (4.9) имеют нули, наличие которых обусловлено осциллирующим характером передаточных функций. Поскольку спектр искаженного изображения равен произведению спектра исходного изображения и передаточной функции [см. (4.5)], наличие нулей приводит к полной утрате данных об исходном изображении на соответствующих частотах. По этой причине не удастся абсолютно точно восста-

новить исходное изображение по наблюдаемому изображению, даже если отсутствуют шумы наблюдения и размеры кадров неограниченны.

При решении задач восстановления изображений используются различные алгоритмы — как имеющие строгое математическое обоснование, так и эвристические. Для искажений, описываемых уравнением свертки, эти алгоритмы условно можно разделить на три основные группы: алгоритмы решения системы алгебраических уравнений, алгоритмы фильтрации изображений в частотной области и итерационные алгоритмы.

#### 4.2. АЛГЕБРАИЧЕСКИЕ МЕТОДЫ ВОССТАНОВЛЕНИЯ ИЗОБРАЖЕНИЙ

Соотношение (4.4) для цифровых изображений фактически представляет собой систему линейных алгебраических уравнений относительно  $u(i_1, i_2)$ . Поэтому задача восстановления исходного изображения  $u(i_1, i_2)$  при известной ФРТ  $h(i_1, i_2)$  может быть сведена к решению такой системы. Удобно представить соотношения (4.3) и (4.4) в матричной форме, используя лексико-графическое упорядочивание. Для этого двумерный массив наблюдаемого изображения  $s(i_1, i_2)$  размером  $L^{(s)} = (L_1^{(s)} + 1) \times (L_2^{(s)} + 1)$  преобразуем в вектор-столбец с размером  $L^{(s)} \times 1$ . Преобразование осуществляется разверткой массива  $s(i_1, i_2)$  по строкам. Аналогичным образом преобразуются в вектор-столбцы  $\mathbf{z}$ ,  $\mathbf{u}$  и  $\mathbf{n}$  искаженное при отсутствии шума изображение  $z(i_1, i_2)$ , исходное изображение  $u(i_1, i_2)$  и шум  $n(i_1, i_2)$ . Размеры векторов  $\mathbf{z}$ ,  $\mathbf{u}$  и  $\mathbf{n}$  равны  $L^{(u)} = L^{(z)} = L^{(n)} = (L_1^{(u)} + 1) \times (L_2^{(u)} + 1)$ . Связь между лексико-графически упорядоченными изображениями определяется соотношением

$$\mathbf{s} = \mathbf{z} + \mathbf{n}, \quad (4.15)$$

где искаженное изображение

$$\mathbf{z} = \mathbf{h}\mathbf{u}. \quad (4.16)$$

Символ  $\mathbf{h}$  обозначает прямоугольную матрицу размером  $L^{(z)} \times L^{(u)}$ , с помощью которой вектор исходного изображения  $\mathbf{u}$  преобразуется в искаженное изображение  $\mathbf{z}$ . Матрица  $\mathbf{h}$  имеет

блочную структуру [4.2], элементы которой представляют собой отсчеты ФРТ. Задачи восстановления изображений алгебраическими методами при наличии и отсутствии шумов наблюдения имеют качественные различия.

Если шумами наблюдения можно пренебречь, то задача восстановления изображения сводится к нахождению оценки (решения)  $\hat{\mathbf{u}}$  матричного уравнения (4.16), удовлетворяющей условию

$$\mathbf{z} = \mathbf{h} \hat{\mathbf{u}}. \quad (4.17)$$

Если бы  $\mathbf{h}$  была квадратной матрицей и существовала обратная матрица  $(\mathbf{h})^{-1}$ , то очевидно, что решение системы имело бы вид

$$\hat{\mathbf{u}} = (\mathbf{h})^{-1} \mathbf{z}. \quad (4.18)$$

Однако матричное уравнение (4.16) представляет собой *недоопределенную* систему линейных алгебраических уравнений, так как количество неизвестных  $L^{(u)}$  больше числа уравнений  $L^{(z)}$  (размеры исходного изображения всегда больше размеров искаженного изображения). Поэтому матрица  $\mathbf{h}$  является прямоугольной матрицей размером  $L^{(z)} \times L^{(u)}$ . В этом случае для отыскания решения используются различные методы псевдообращения матриц, которые описаны в [4.3]. Если недоопределенная система (4.16) разрешима, то она имеет несколько решений. Возникает проблема выбора единственного решения из множества возможных, которое и будет принято за оценку  $\hat{\mathbf{u}}$ . Среди всех возможных решений недоопределенной разрешимой системы (4.16) в качестве оценки  $\hat{\mathbf{u}}$  выбирается решение, минимизирующее норму ошибки восстановления

$$\varepsilon_1^2 = \mathbf{e}_1(\hat{\mathbf{u}})^T \mathbf{e}_1(\hat{\mathbf{u}}) = \sum_{i=1}^{L^{(u)}} (u_i - \hat{u}_i)^2 = \min_{\mathbf{u}} , \quad (4.19)$$

где  $T$  — символ транспонирования;  $\mathbf{e}_1(\hat{\mathbf{u}}) = \mathbf{u} - \hat{\mathbf{u}}$  — вектор ошибки восстановления. Критерий (4.19) называется критерием наименьших квадратов. Доказано [4.3], что норма ошибки будет минимальной, если оценка

$$\hat{\mathbf{u}} = \mathbf{h}^+ \mathbf{z}, \quad (4.20)$$

где  $\mathbf{h}^+ = \mathbf{h}^T (\mathbf{h} \mathbf{h}^T)^{-1}$  — обобщенная обратная матрица. В общем случае норма ошибки не равна нулю.

Точное восстановление исходного изображения при отсутствии шумов возможно, во-первых, когда искаженное изображение получено в результате циклической свертки исходного изображения и ФРТ; во-вторых, когда объекты исходного изображения расположены в центре кадра и наблюдаются на фоне постоянной яркости, причем расстояние от объектов до границ кадра больше апертуры ФРТ. В том и другом случае число неизвестных будет равно числу уравнений, так как объекты, расположенные вне кадра, не будут влиять на яркость наблюдаемого изображения. Иными словами, точное восстановление при отсутствии шума возможно тогда, когда ограничение размеров кадра наблюдаемого изображения не приводит к потере информации об исходном изображении.

Для искаженных изображений, наблюдаемых в присутствии шумов, к элементам вектора-столбца  $\mathbf{z}$  добавляются отсчеты вектора-столбца  $\mathbf{n}$ . Это делает систему уравнений, как правило, неразрешимой. Неразрешимость системы означает, что не существует оценки исходного изображения, при которой она перейдет в тождество. Можно найти лишь приближенное решение неразрешимой системы, которое определяется из условия минимума нормы ошибки [4.4, 4.5]

$$\varepsilon_2^2 = (\mathbf{e}_2(\hat{\mathbf{u}}))^T \mathbf{e}_2(\hat{\mathbf{u}}) = (\mathbf{s} - \mathbf{h}\hat{\mathbf{u}})^T (\mathbf{s} - \mathbf{h}\hat{\mathbf{u}}) = \min_{\hat{\mathbf{u}}} . \quad (4.21)$$

Здесь оптимальным оператором [в смысле критерия наименьших квадратов (4.21)], формирующим оценку  $\hat{\mathbf{u}}$ , также является обобщенная обратная матрица  $\mathbf{h}^+$ . Причем этот оператор — единственный оператор, обеспечивающий минимум нормы

$$\text{оценки } (\hat{\mathbf{u}})^T \hat{\mathbf{u}} = \sum_{i=1}^{L^{(s)}} \hat{u}_i^2 .$$

Таким образом, в обоих рассмотренных случаях обобщенное обращение матриц дает оптимальное решение, удовлетворяющее критериям наименьших квадратов (4.19) или (4.21). Следует подчеркнуть, что, несмотря на одинаковые названия, по сути это два разных критерия. Для разрешимой недоопределенной системы (4.16) (когда выбирается одно решение из множества возможных) ошибка  $\mathbf{e}_2(\hat{\mathbf{u}})$  равна нулю. В противном случае ошибка  $\mathbf{e}_2(\hat{\mathbf{u}})$  всегда отлична от нуля, так как точное решение системы отсутствует.



Основным недостатком алгебраических алгоритмов восстановления изображений является необходимость выполнения трудоемких операций обращения, умножения и транспонирования матриц огромных размеров. Напомним, что размер матрицы  $\mathbf{h}$  равен произведению числа отсчетов исходного и наблюдаемого изображений. Кроме того, обращение матриц больших размеров представляет собой трудную задачу. Эта задача значительно упрощается, если искаженное изображение формируется из исходного путем циклической свертки с ФРТ. К сожалению, для реальных задач восстановления изображений это условие не выполняется. Альтернативой алгебраическим методам являются методы линейной фильтрации изображений.

#### 4.3. МЕТОДЫ ВОССТАНОВЛЕНИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ПРОСТРАНСТВЕННОЙ ФИЛЬТРАЦИИ

Методы восстановления изображений, которые будут рассмотрены в данном разделе, реализуются с помощью ДПФ в частотной области. При этом обычная свертка заменяется циклической как в модели формирования искаженного изображения (4.4), так и в процедуре восстановления методом пространственной фильтрации. Все изображения  $\tilde{s}(i_1, i_2)$ ,  $\tilde{z}(i_1, i_2)$ ,  $\tilde{h}(i_1, i_2)$ ,  $\tilde{u}(i_1, i_2)$  и ФРТ полагаются периодически продолженными и имеют одинаковые размеры кадра  $L = L_1 \times L_2$ . Замена обычной свертки циклической в модели формирования приводит к тому, что при синтезе восстанавливающего фильтра не учитывается факт влияния объектов, расположенных вне поля зрения объекта, на значение яркости на краях искаженного изображения. При восстановлении реально искаженных изображений таким фильтром возникают краевые эффекты, компенсация которых является одной из основных задач при реализации алгоритмов восстановления изображения на основе пространственной фильтрации.

При циклической свертке модель (4.3), (4.4) формирования искаженного изображения определяется соотношением

$$\begin{aligned} \tilde{s}(i_1, i_2) &= \tilde{z}(i_1, i_2) + \tilde{h}(i_1, i_2) = \\ &= \tilde{h}(i_1, i_2) \otimes \tilde{u}(i_1, i_2) + \tilde{h}(i_1, i_2), \quad (i_1, i_2) \in \Theta, \end{aligned} \quad (4.22)$$

где  $\Theta$  — кадр изображения, одинаковый для всех изображений и ФРТ, входящих в (4.22). Размеры кадра равны периоду повторения изображений и ФРТ.

Применяя к (4.22) ДПФ, получаем

$$\begin{aligned}\tilde{S}(k_1, k_2) &= \tilde{Z}(k_1, k_2) + \tilde{N}(k_1, k_2) = \\ &= \tilde{H}(k_1, k_2)\tilde{U}(k_1, k_2) + \tilde{N}(k_1, k_2), \quad (k_1, k_2) \in \Theta.\end{aligned}\quad (4.23)$$

Система восстановления изображений на основе пространственной фильтрации представляет собой линейный пространственно-инвариантный двумерный фильтр. На выходе этого фильтра формируется оценка

$$\hat{u}(i_1, i_2) = \tilde{h}_s(i_1, i_2) \otimes \tilde{s}(i_1, i_2) \quad (4.24)$$

исходного изображения  $u(i_1, i_2)$ . В пространственно-частотной области спектр оценки с учетом (4.24) можно записать как

$$\begin{aligned}\hat{U}(k_1, k_2) &= \tilde{H}_s(k_1, k_2)\tilde{S}(k_1, k_2) = \\ &= \tilde{H}_s(k_1, k_2)(\tilde{H}(k_1, k_2)\tilde{U}(k_1, k_2) + \tilde{N}(k_1, k_2)).\end{aligned}\quad (4.25)$$

#### 4.3.1. ИНВЕРСНЫЙ ФИЛЬТР

Простейшим способом восстановления четкости изображения является обработка наблюдаемого изображения в пространственно-частотной области инверсным фильтром [4.5]. Передаточная функция инверсного восстанавливающего фильтра определяется соотношением

$$\tilde{H}_{\text{инв}}(k_1, k_2) = 1 / \tilde{H}(k_1, k_2). \quad (4.26)$$

Она выбирается из условия  $\tilde{H}_{\text{инв}}(k_1, k_2)\tilde{H}(k_1, k_2) = 1$ , обеспечивающего компенсацию искажений, вносимых ФРТ формирующей системы. При этом спектр оценки исходного изображения

$$\hat{U}(k_1, k_2) = \tilde{U}(k_1, k_2) + \tilde{N}(k_1, k_2) / \tilde{H}(k_1, k_2). \quad (4.27)$$

Таким образом, восстановленное изображение равно сумме исходного изображения и шума наблюдения, прошедшего через инверсный фильтр. При отсутствии шума достигается точное восстановление инверсным фильтром исходного изображения  $\hat{u}(i_1, i_2)$  по искаженному изображению  $\tilde{z}(i_1, i_2)$ . При восстановлении изображений инверсным фильтром возникают краевые эффекты, которые проявляются в виде осциллирующей помехи большой мощности, полностью маскирующей восстановленное

изображение. Краевые эффекты возникают даже при отсутствии шума наблюдения.

На рис. 4.13 и 4.14 приведены результаты восстановления изображений «Часы» и «Сатурн» инверсным фильтром: а) исходные изображения размером  $200 \times 200$  элементов; б) дефокусированные изображения, полученные в результате свертки с гауссовским импульсом при  $\alpha = 0.1$  с последующим «обрезанием» краев до размеров  $170 \times 170$  элементов; в) изображения, восстановленные инверсным фильтром. Восстановить изображение «Часы» инверсным фильтром не удастся из-за краевых эффектов. Практически идеальное восстановление изображения «Сатурн» объясняется тем, что объекты наблюдаются на фоне постоянной яркости и расположены в центре кадра. В этом случае изображения  $z(i_1, i_2)$  и  $\tilde{z}(i_1, i_2)$ , полученные в результате обычной и циклической свертки с ФРТ, равны друг другу. Отметим, что при этих условиях алгебраический метод также позволяет точно восстановить изображение. Однако при инверсной фильтрации процедура обращения матриц заменяется на более простую процедуру перемножения массивов в частотной области.

На рис. 4.15 и 4.16 показаны сечения типичных частотных характеристик ФРТ и соответствующих им инверсных фильтров. Из приведенных зависимостей следует, что модуль передаточной функции формирующей системы, как правило, стремится к нулю на высоких частотах. Кроме того, нули в передаточной функции имеются в рабочей полосе частот при расфокусировке камеры (4.10) и смазе (4.6). В этом случае инверсный фильтр является сингулярным, так как модуль его передаточной функции становится бесконечно большим на пространственных частотах, соответствующих нулевым значениям модуля передаточной функции искажающей системы. Причем наличие даже относительно слабого шума приводит к появлению интенсивных шумовых составляющих на выходе инверсного фильтра, полностью разрушающих изображение. Этот факт иллюстрирует рис. 4.17. К дефокусированному изображению «Сатурн» (см. рис. 4.14, б) был добавлен аддитивный дельта-коррелированный шум. Из восстановленного изображения видно, что даже при пренебрежимо малом уровне шума

$\left( \text{отношение сигнал/шум } q^2 = \frac{D_z}{D_n} = 10^{27} \right)$  метод инверсной

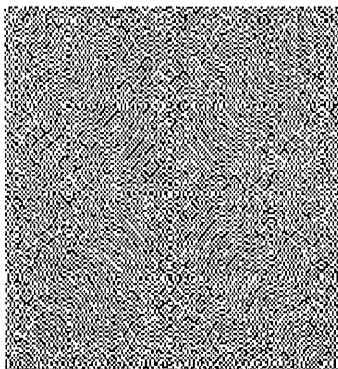
фильтрации дает очень плохой результат.



а

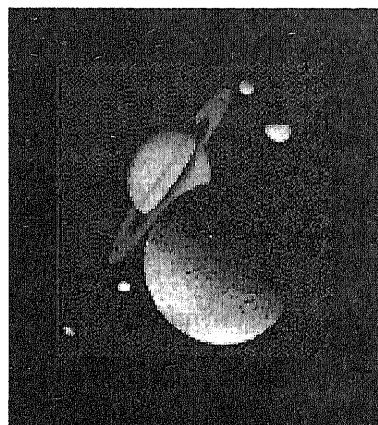


б



в

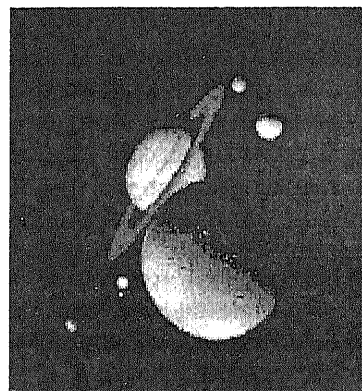
Рис. 4.15. Результаты восстановления изображения «Часы»



а



б



в

Рис. 4.14. Результаты восстановления изображения «Сатурн»

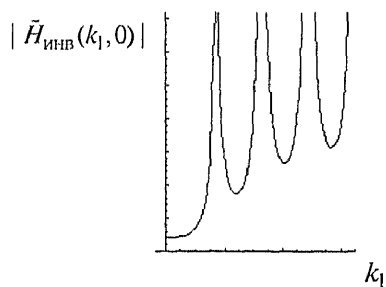
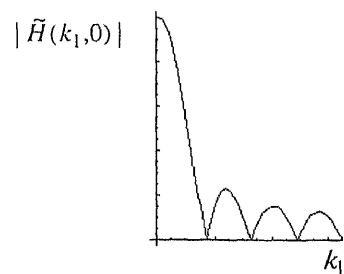


Рис. 4.15. Частотные характеристики искажающей системы с цилиндрической ФРТ и инверсного фильтра

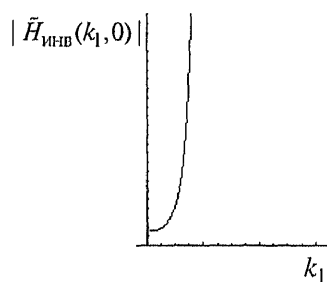
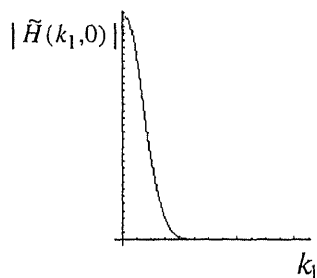


Рис. 4.16. Частотные характеристики искажающей системы с гауссовской ФРТ и инверсного фильтра

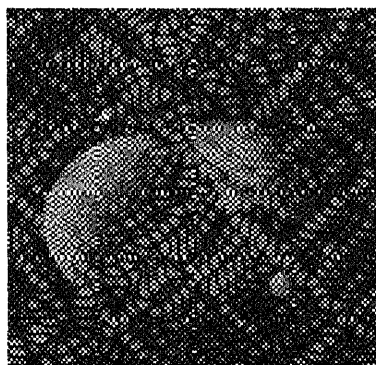


Рис. 4.17. Результат восстановления изображения «Сатурн» при  $q^2=10^{27}$

Существуют частные методы ослабления шумов, которые заключаются в ограничении полосы инверсного фильтра. Последовательно с инверсным фильтром включают корректирующее звено, модуль передаточной функции которого стремится к нулю за пределами некоторой наперед заданной граничной частоты. При этом граничную частоту выбирают из компромисса между снижением уровня шума и четкостью восстановленного изображения. Однако эти методы не решают проблем краевых эффектов и наличия нулей передаточной функции формирующей системы в рабочем диапазоне частот.

Таким образом, несмотря на очевидную простоту метода инверсной фильтрации, он может успешно использоваться для восстановления ограниченного класса изображений, у которых уровень фона на краях постоянен. Однако метод инверсной фильтрации обладает чрезвычайно низкой помехоустойчивостью.

#### 4.3.2. ФИЛЬТР ВИНЕРА

Инверсная фильтрация обладает низкой помехоустойчивостью, потому что этот метод не учитывает зашумленность наблюдаемого изображения. Значительно менее подвержен влиянию помех и сингулярностей, обусловленных нулями передаточной функции искажающей системы, фильтр Винера (смотри главу 3), так как при его синтезе наряду с видом ФРТ используется информация о спектральных плотностях мощности изображения и шума. При этом полагается, что изображение является реализацией случайного двумерного поля. Частотная характеристика восстанавливающего фильтра Винера, полученная для периодически продолженных изображений, с учетом (2.34) имеет вид [4.6]

$$\tilde{H}_{\text{Вин}}(k_1, k_2) = \frac{\tilde{G}_{u,s}(k_1, k_2)}{\tilde{G}_s(k_1, k_2)} = \frac{\tilde{H}^*(k_1, k_2) \tilde{G}_u(k_1, k_2)}{|\tilde{H}(k_1, k_2)|^2 \tilde{G}_u(k_1, k_2) + \tilde{G}_n(k_1, k_2)}, \quad (4.28)$$

где  $\tilde{G}_n(k_1, k_2)$ ,  $\tilde{G}_s(k_1, k_2)$ ,  $\tilde{G}_u(k_1, k_2)$  — спектральные плотности мощности периодически продолженных шума, наблюдаемого и исходного изображений;  $\tilde{G}_{u,s}(k_1, k_2)$  — взаимная спектральная плотность мощности периодически продолженных исходного и наблюдаемого изображений;  $(\cdot)^*$  — символ комплексного сопряжения. Как и при инверсной фильтрации, обработка производится в частотной области.

Преобразуем передаточную функцию фильтра Винера (4.28) следующим образом:

$$\tilde{H}_{\text{Вин}}(k_1, k_2) = \frac{\tilde{H}^*(k_1, k_2)}{|\tilde{H}(k_1, k_2)|^2 + \frac{\tilde{G}_n(k_1, k_2)}{\tilde{G}_u(k_1, k_2)}}. \quad (4.29)$$

Анализируя соотношения (4.28) и (4.29), можно отметить следующее:

1. При отсутствии шума фильтр Винера переходит в инверсный фильтр. Следовательно, в области низких частот, где, как правило, отношение сигнал/шум велико, передаточные функции инверсного и винеровского фильтров практически совпадают.

2. При уменьшении спектральной плотности мощности исходного изображения передаточная функция фильтра Винера стремится к нулю. Для изображений это характерно на верхних частотах.

3. На частотах, соответствующих нулям передаточной функции формирующей системы, передаточная функция фильтра Винера также равна нулю. Таким образом, решается проблема сингулярности восстанавливающего фильтра.

На рис. 4.18 приведены одномерные сечения типичных передаточных функций винеровских фильтров (сплошная линия). Здесь же для сравнения приведены сечения передаточных функций инверсных фильтров (4.15) и (4.16), которые обозначены штриховой линией.

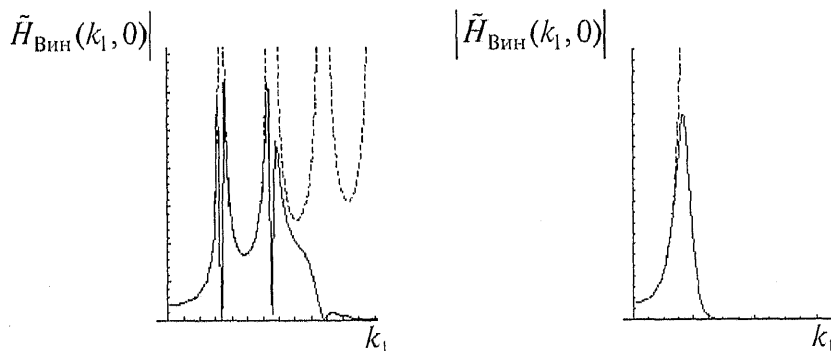


Рис. 4.18. Частотные характеристики фильтра Винера при цилиндрической и гауссовской ФРТ

Рассмотрим результаты моделирования винеровского алгоритма восстановления. На рис. 4.19,*а* и 4.21,*а* приведены результаты искажения изображений «Сатурн» и «Часы» сверткой с гауссовской ФРТ ( $\alpha = 0.1$ ) с последующим «обрезанием» краев и добавлением аддитивного дельта-коррелированного шума ( $q^2 = 10^4$ ). На рис. 4.20,*а* и 4.22,*б* приведены изображения, полученные в результате смаза ( $L_{i_2}^{(h)} = 15$ ) изображений «Сатурн» (рис. 4.6) и «Часы» (рис. 4.22,*а*) ( $L_{i_2}^{(h)} = 15$ ) также с последующим «обрезанием» краев и добавлением аддитивного дельта-коррелированного шума ( $q^2 = 10^4$ ).

Размеры всех наблюдаемых и восстановленных изображений равны  $170 \times 170$  элементов. Результаты восстановления винеровским фильтром изображения «Сатурн» (рис. 4.19,*б* и 4.20,*б*) свидетельствуют о том, что фильтр Винера значительно лучше подавляет шумы. Осциллирующая помеха на результатах восстановления изображения «Часы» (рис. 4.21,*б* и 4.22,*в*) вызвана краевыми эффектами. Очевидно, что ее уровень существенно меньше, чем при инверсной фильтрации (см. рис. 4.13,*в*). Однако винеровский фильтр лишь частично компенсирует краевые эффекты, которые делают качество восстановления неудовлетворительным.

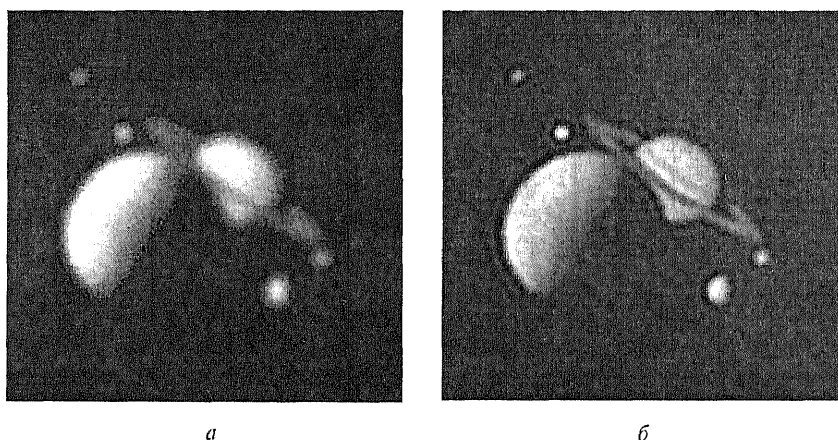


Рис. 4.19. Восстановление дефокусированного изображения «Сатурн»  
при  $q^2 = 10^4$



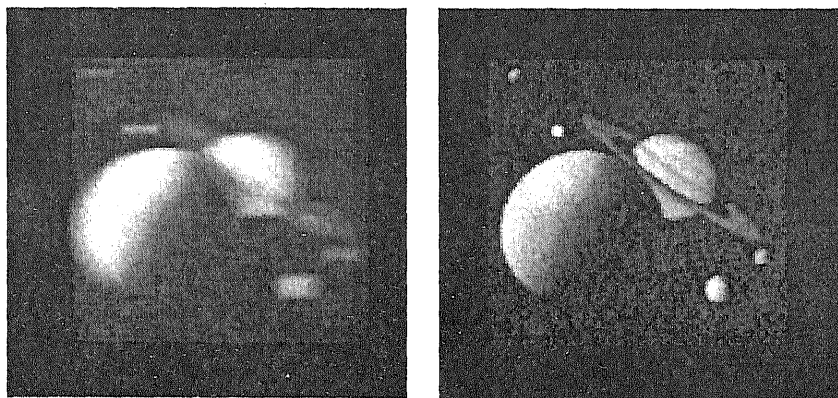


Рис. 4.20. Восстановление смазанного изображения «Сатурн»  
при  $q^2 = 10^4$

Таким образом, за счет использования информации о спектральных характеристиках изображения и шума фильтр Винера обладает относительно высокой помехоустойчивостью и у него отсутствует сингулярность, обусловленная нулями передаточной функции формирующей системы. Основным недостатком фильтра Винера остается наличие краевых эффектов, которые проявляются в виде осциллирующей помехи, маскирующей восстановленное изображение.

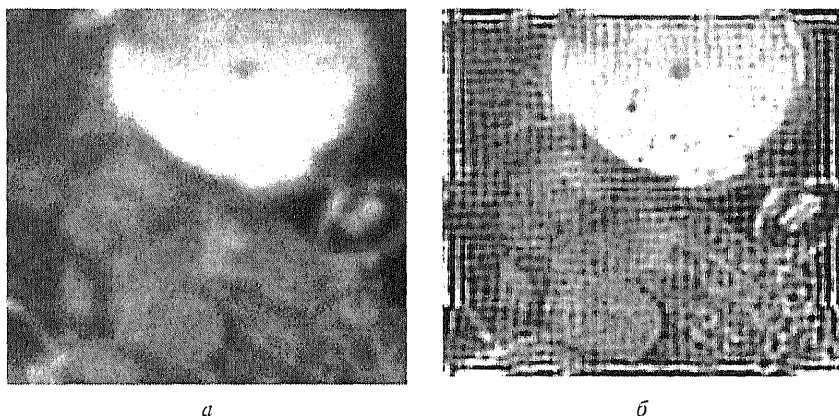


Рис. 4.21. Восстановление дефокусированного изображения «Часы»  
при  $q^2 = 10^4$

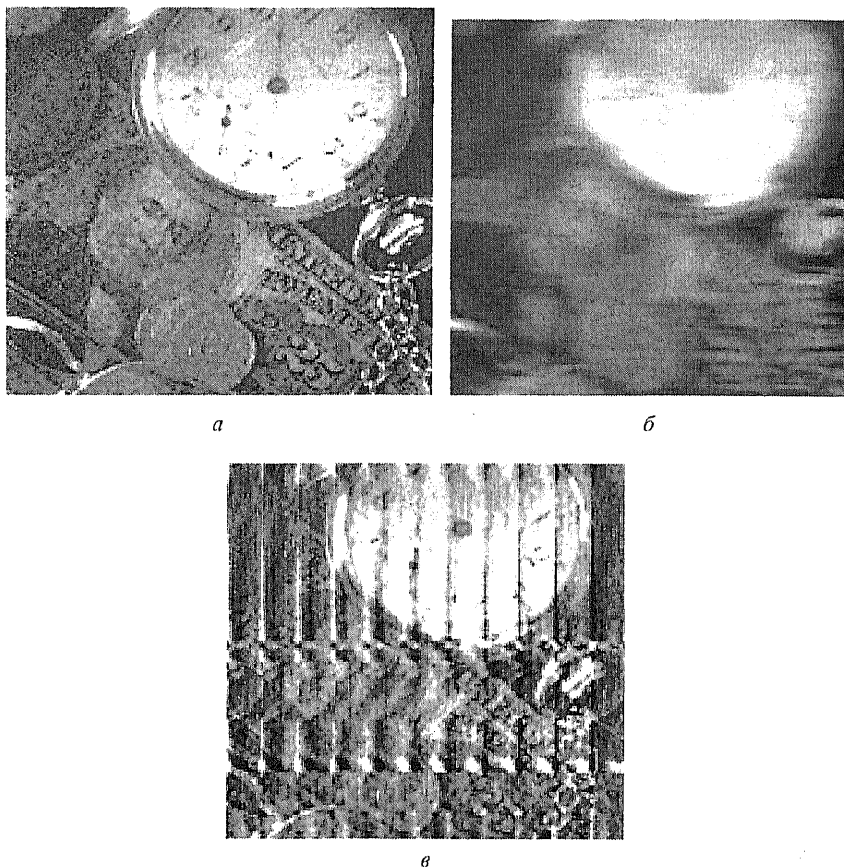


Рис. 4.22. Восстановление смазанного изображения «Часы»  
при  $q^2 = 10^4$

#### 4.3.3. КОМПЕНСАЦИЯ КРАЕВЫХ ЭФФЕКТОВ ПРИ ВОССТАНОВЛЕНИИ ЛИНЕЙНО ИСКАЖЕННЫХ ИЗОБРАЖЕНИЙ

На восстановленных изображениях, приведенных на рис. 4.13, в, 4.21, б и 4.22, в, присутствует осциллирующая помеха большой интенсивности, которая возникает из-за того, что инверсный фильтр и фильтр Винера были синтезированы без учета ограниченных размеров наблюдаемых изображений. Вследствие того, что искаженное изображение записывается в кадре конечного размера, в усеченном изображении происходит потеря

информации, содержащейся в исходном изображении вблизи границ. Поэтому при коррекции линейных искажений усеченного изображения возникают ложные детали в виде ряби или полос, интенсивность которых особенно велика при цилиндрической форме ФРТ и равномерном смазе.

К сожалению, решить уравнение Винера-Хопфа для сигналов и изображений, наблюдаемых на ограниченном интервале, не удастся. Поэтому отсутствуют оптимальные пространственно-инвариантные фильтры, учитывающие краевые эффекты. Для компенсации краевых эффектов используются различные эвристические алгоритмы. Некоторые из них будут рассмотрены в данном подразделе.

Если нас интересует центральная часть изображения и его размеры значительно больше размеров кадра ФРТ, то для компенсации краевых эффектов применяют умножение наблюдаемого изображения на функцию окна  $w(i_1, i_1)$ , которая плавно уменьшается до нуля на краях кадра  $\Theta^{(s)}$  наблюдаемого изображения и равна нулю всюду за его пределами. После этого изображение восстанавливается фильтром Винера.

При дефокусировке функция окна полагается разделимой относительно пространственных координат:  $w(i_1, i_2) = w(i_1)w(i_2)$ . При вертикальном или горизонтальном смазе используется одномерная функция окна, на которую умножаются соответственно столбцы или строки наблюдаемого изображения. Известно много одномерных функций окна  $w(i)$ , которые могут быть использованы при восстановлении изображений, например, окна Бартлетта, Кайзера, Блэкмана и др [4.7].

Хорошие результаты дает функция окна [4.8]

$$w_r(i | \gamma, \beta) = 0.5 \left( \operatorname{th} \left( \left( i + \frac{\gamma}{2} \right) / \beta \right) - \operatorname{th} \left( \left( i - \frac{\gamma}{2} \right) / \beta \right) \right), \quad (4.30)$$

форма которой определяется двумя независимыми параметрами  $\gamma$  и  $\beta$ . Параметр  $\gamma$  влияет на размеры окна, а  $\beta$  — на скорость спада краев окна к нулю.

На рис. 4.23 и 4.24 приведены результаты восстановления изображения «Часы» при горизонтальном смазе, где  $a$  — результаты умножения строк искаженного изображения, приведенного на рис. 4.22,  $b$ , на окно Кайзера и окно (4.30);  $b$  — результаты восстановления фильтром Винера. Параметры окон подбирались, исходя из визуального качества восстанавливаемых изображений.

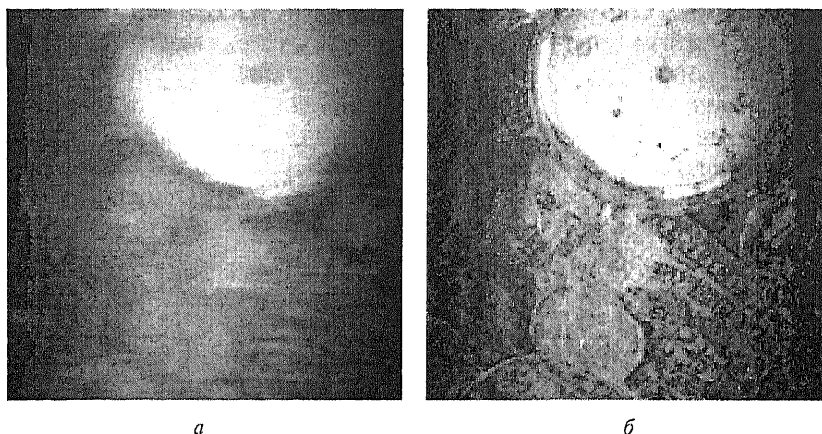


Рис. 4.23. Восстановление с использованием окна Кайзера

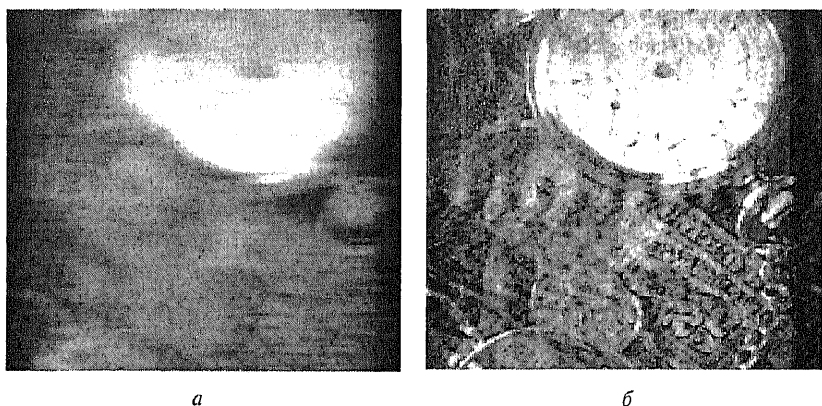


Рис. 4.24. Восстановление с использованием окна (4.30)

Уровень яркости на краях изображений, умноженных на окно, стремится к нулю, поэтому вместе с уменьшением краевых эффектов сужаются границы восстанавливаемого изображения. Кроме того, оптимальные параметры окон зависят от параметров искажающей системы и определяются опытным путем, что затрудняет практическое применение алгоритмов восстановления.

Учесть ограниченные размеры наблюдаемого изображения можно на этапе синтеза фильтра Винера, который использует информацию о спектрально-корреляционных характеристиках изображения. Получение изображения ограниченных размеров эквивалентно умножению бесконечного изображения на окно

единичной яркости, размеры которого равны размерам кадра  $\Theta^{(s)}$ . Очевидно, что спектрально-корреляционные характеристики такого усеченного изображения будут отличаться от аналогичных характеристик бесконечных изображений. Корреляционная функция усеченного изображения может быть получена путем умножения корреляционной функции неограниченного изображения на окно [4.9]:

$$\psi(i_1, i_2) = \begin{cases} \prod_{j=1}^2 \left( 1 - \frac{|i_j|}{L_j^{(s)} + 1} \right), & \text{при } |i_j| \leq L_j^{(s)}, \\ 0, & \text{при } |i_j| > L_j^{(s)}. \end{cases} \quad (4.31)$$

В этом случае спектральная плотность мощности усеченного изображения равна свертке спектральных плотностей мощности неограниченного изображения и окна (4.31). Подставляя соответствующие спектральные плотности мощности в уравнение Винера-Хинчина и решая его, получаем коэффициент передачи фильтра для усеченного изображения [4.10]:

$$H_{\text{Вин}}(\omega_1, \omega_2) = \frac{G_{u,s}(\omega_1, \omega_2) \otimes \Psi(\omega_1, \omega_2)}{G_s(\omega_1, \omega_2) \otimes \Psi(\omega_1, \omega_2)}, \quad (4.32)$$

где  $\Psi(\omega_1, \omega_2)$  — спектральная плотность окна (4.31). Следует подчеркнуть, что импульсная характеристика фильтра (4.32) не сводится к произведению импульсной характеристики фильтра Винера и regularизирующего двумерного треугольного окна (4.31).

На рис. 4.25 приведен результат восстановления изображения «Часы» фильтром (4.32), откуда следует, что фильтр (4.32) практически полностью компенсирует краевые эффекты. Это позволяет отказаться от предварительной обработки. Качество восстановления изображения в центре и на краях почти одинаковое. Параметры фильтра (4.32) полностью определяются исходными данными и не требуют дополнительной подстройки. При использовании быстрого преобразования Фурье

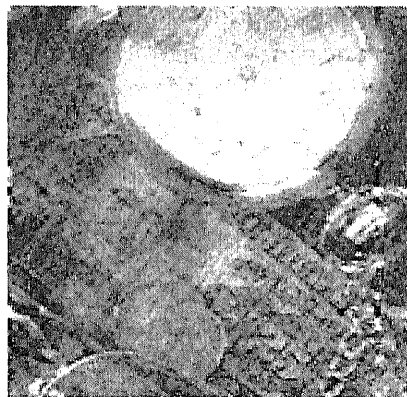


Рис. 4.25. Результат восстановления с компенсацией краевых эффектов

для обработки изображений объем вычислений при реализации фильтра (4.32) такой же, как и для фильтра Винера (4.28).

Иногда наряду с компенсацией краевых эффектов требуется расширить границы восстанавливаемого изображения, чтобы извлечь больший объем информации об исходном изображении. Для этого используется процедура экстраполяции, которая состоит в том, что двумерную функцию яркости  $s(i_1, i_2)$  наблюдаемого изображения продолжают с границ кадра  $\Theta^{(s)}$  на кадр больших размеров  $\Theta^{(a)}$  так, чтобы функция яркости была гладкой и на границах кадра  $\Theta^{(a)}$  равнялась нулю.

Простейшей является процедура одномерной экстраполяции. Яркость изображения вдоль строк и столбцов за пределами кадра наблюдаемого изображения  $\Theta^{(s)}$  задается в виде полинома

$$f(i_j) = a_0 + a_1 i_j + a_2 i_j^2 \dots, \quad j = \overline{1, 2}.$$

Коэффициенты  $\{a_k\}$  определяются исходя из требований, предъявляемых к свойствам функции яркости. Например, на границе кадра  $\Theta^{(a)}$  функция яркости должна равняться нулю, она должна быть неотрицательной, максимальное значение экстраполирующей функции не должно превышать максимального значения наблюдаемого изображения и т.п. Метод экстраполяции иллюстрируется рис. 4.26, где  $a$  — экстраполированное изображение;  $b$  — результат восстановления. Размер наблюдаемого изображения «Часы» (см. рис. 4.22,  $b$ ) равен  $170 \times 170$  элементов, экстраполированного —  $200 \times 200$  элементов. В качестве экстраполирующей функции использован полином первой степени.

Следует обратить внимание на то, что применение процедуры экстраполяции позволило восстановить изображение в пределах кадра исходного изображения, размеры которого больше размеров кадра наблюдаемого изображения на величину смаза. Очевидно, что в центральной части качество восстановления при экстраполяции хуже, чем при умножении на окно. Однако эта процедура позволяет увеличить размеры кадра восстановленного изображения.

Улучшить качество восстановления можно, используя одновременно экстраполяцию наблюдаемого изображения и фильтра (4.32). Результат такой комбинированной процедуры приведен на рис. 4.27. Рассмотренные методы восстановления являются линейными. Их широкое использование обусловлено достаточно простыми методами синтеза и анализа линейных систем, а также высокой вычислительной эффективностью. Однако эти методы

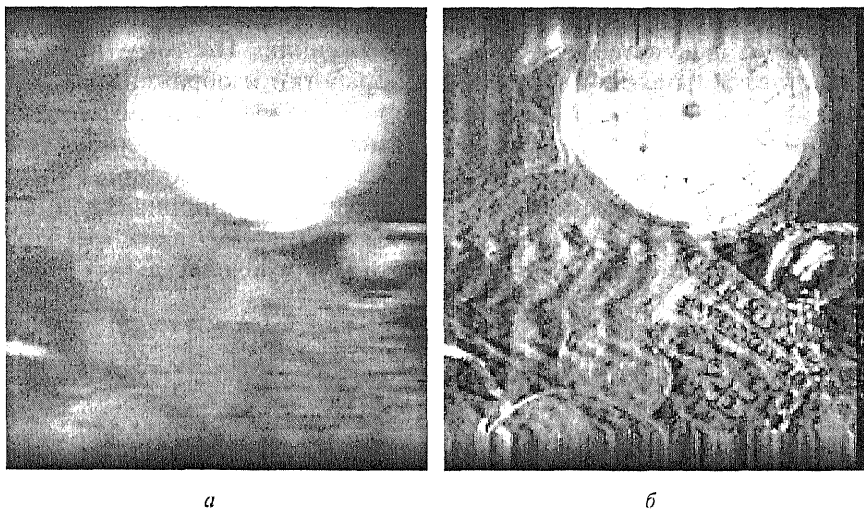


Рис. 4.26. Восстановление с применением экстраполяции

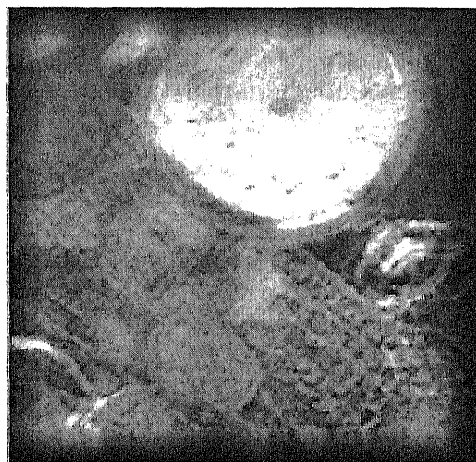


Рис. 4.27. Восстановление с использованием экстраполяции и компенсации краевых эффектов

не оптимальны и не всегда обеспечивают эффективную компенсацию искажений. Линейная обработка служит лишь приближением к оптимальной обработке, так как статистические характеристики подавляющего большинства изображений негауссов-

ские. Кроме того, линейные методы не учитывают априорные данные о восстанавливаемых изображениях. Поэтому интерес представляют нелинейные методы обработки изображений. Синтез оптимальных нелинейных алгоритмов, как правило, значительно сложнее, чем линейных. Однако существуют линейные методы восстановления, которые достаточно просто могут быть преобразованы в нелинейные, учитывающие априорные данные об изображениях и помехах. Ярким примером таких методов служат итерационные методы (методы последовательных приближений).

#### 4.4. ИТЕРАЦИОННЫЕ МЕТОДЫ ВОССТАНОВЛЕНИЯ ИЗОБРАЖЕНИЙ

Итерационными методами называют способы решения задач, в которых, выбирая некоторое начальное приближенное решение, вычисляют следующие, более точные приближения, используя предыдущие.

Рассмотрим один из способов построения итерационных процедур, основанный на разложении в ряд частотной характеристики инверсного фильтра [4.6]. Спектр оценки исходного изображения при инверсной фильтрации определяется соотношением

$$\widehat{U}(\omega_1, \omega_2) = \frac{1}{H(\omega_1, \omega_2)} S(\omega_1, \omega_2). \quad (4.33)$$

Представим передаточную функцию инверсного фильтра  $\frac{1}{H(\omega_1, \omega_2)}$  в виде геометрической прогрессии:

$$\frac{1}{H(\omega_1, \omega_2)} = \sum_{l=0}^{\infty} (1 - H(\omega_1, \omega_2))^l. \quad (4.34)$$

Подставив (4.34) в (4.33), получим

$$\begin{aligned} \widehat{U}(\omega_1, \omega_2) = & S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))S(\omega_1, \omega_2) + \\ & + (1 - H(\omega_1, \omega_2))^2 S(\omega_1, \omega_2) + \dots \end{aligned} \quad (4.35)$$

Соотношение (4.35) позволяет представить процедуру нахождения оценки  $\widehat{U}(\omega_1, \omega_2)$  в виде последовательных приближений:



$$\widehat{U}^{(0)}(\omega_1, \omega_2) = S(\omega_1, \omega_2); \quad (4.36)$$

$$\begin{aligned} \widehat{U}^{(1)}(\omega_1, \omega_2) &= S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))S(\omega_1, \omega_2) = \\ &= S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))\widehat{U}^{(0)}(\omega_1, \omega_2); \end{aligned}$$

$$\begin{aligned} \widehat{U}^{(2)}(\omega_1, \omega_2) &= S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))S(\omega_1, \omega_2) + \\ &+ (1 - H(\omega_1, \omega_2))^2 S(\omega_1, \omega_2) = \\ &= S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))(S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))^2 S(\omega_1, \omega_2)) = \\ &= S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))\widehat{U}^{(1)}(\omega_1, \omega_2); \end{aligned}$$

.....

$$\widehat{U}^{(n)}(\omega_1, \omega_2) = S(\omega_1, \omega_2) + (1 - H(\omega_1, \omega_2))\widehat{U}^{(n-1)}(\omega_1, \omega_2),$$

где каждое последующее приближение вычисляется по предыдущему. Взяв преобразование Фурье от соотношений (4.36), получим итерационную процедуру Ван Циттерта [4.11]:

$$\begin{aligned} \widehat{u}^{(0)}(i_1, i_2) &= s(i_1, i_2); \\ \widehat{u}^{(n)}(i_1, i_2) &= s(i_1, i_2) + (\delta(i_1, i_2) - h(i_1, i_2)) \otimes \otimes \widehat{u}^{(n-1)}(i_1, i_2), \end{aligned} \quad (4.37)$$

которую можно интерпретировать как процедуру последовательного нахождения поправок  $\sigma^{(n)} = (\delta(i_1, i_2) - h(i_1, i_2)) \otimes \otimes \widehat{u}^{(n-1)}(i_1, i_2)$  к искаженному изображению  $s(i_1, i_2)$ . Если в результате последовательных приближений на  $n$ -м шаге будет найдено точное решение:  $\widehat{u}^{(n)}(i_1, i_2) = u(i_1, i_2)$ , то на последующих шагах, как нетрудно убедиться, оценка изменяться не будет. В итерационном алгоритме (4.37) нахождение обратного оператора заменяется на многократное вычисление свертки.

При использовании итерационных алгоритмов необходимо знать ответы на два вопроса: сходится ли он и если сходится, то к какому решению. Сходимость алгоритма (4.37) к решению (4.33) определяется сходимостью ряда бесконечной геометрической прогрессии (4.34). Этот ряд сходится при  $|1 - H(\omega_1, \omega_2)| < 1$ , т.е. когда передаточная функция искажающей системы удовлетворяет условию

$$0 < H(\omega_1, \omega_2) < 1. \quad (4.38)$$

Условие (4.38) выполняется для гауссовской ФРТ. При цилиндрической ФРТ и равномерном смазе соотношение (4.33) заменяют на эквивалентное соотношение

$$\hat{U}(\omega_1, \omega_2) = \frac{1}{|H(\omega_1, \omega_2)|^2} H^*(\omega_1, \omega_2) S(\omega_1, \omega_2).$$

Тогда итерационный алгоритм (4.37) имеет вид [4.6]

$$\hat{u}^{(0)}(i_1, i_2) = h_1(i_1, i_2) \otimes \otimes s(i_1, i_2); \quad (4.39)$$

$$\hat{u}^{(n)}(i_1, i_2) = h_1(i_1, i_2) \otimes \otimes s(i_1, i_2) + (\delta(i_1, i_2) - h_2(i_1, i_2)) \otimes \otimes \hat{u}^{(n-1)}(i_1, i_2),$$

где  $h_1(i_1, i_2)$  и  $h_2(i_1, i_2)$  — импульсные характеристики фильтров с передаточными функциями  $H^*(\omega_1, \omega_2)$  и  $|H(\omega_1, \omega_2)|^2$  соответственно. Свертка в (4.37) и (4.39) может быть выполнена с помощью БПФ в предположении, что изображения и импульсные характеристики являются периодически продолженными.

Очевидно, что рассмотренный итерационный алгоритм является линейным и не имеет никаких преимуществ по сравнению с линейными алгоритмами. Однако этот метод позволяет эффективно бороться с краевыми эффектами и чрезмерным усилением шумов при восстановлении изображений. Итерационный процесс всегда можно остановить, если шум и осциллирующая помеха на изображении резко усиливаются. Остановка итерационного процесса означает усечение ряда (4.34), что приводит к ограничению коэффициента усиления за пределами некоторой граничной частоты. С увеличением

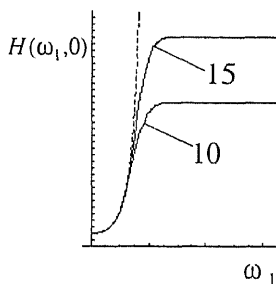


Рис. 4.28. Частотные характеристики итерационного фильтра на разных шагах

длины ряда возрастают граничная частота и коэффициент усиления фильтра. Этот эффект иллюстрируется рис. 4.28, где показаны одномерные сечения частотных характеристик фильтров при 10 и 15 слагаемых в ряде (4.34) (сплошные линии). Здесь же для сравнения приведено одномерное сечение частотной характеристики инверсного фильтра (штриховая линия).

На рис 4.29 показаны результаты восстановления изображения «Часы», где  $a$  и  $b$  — повторно приведенные

исходное (см. рис. 4.22, *а*) и искаженное в результате смаза (рис. 4.22, *б*) изображения; *в* — восстановленное изображение итерационным алгоритмом (4.37) (число итераций  $n = 105$ ); *г* — результат восстановления по экстраполированному наблюдаемому изображению 4.26, *а* (число итераций  $n = 135$ ). В качестве критерия останова итеративного процесса использовался критерий минимума нормированной среднеквадратической ошибки оценивания:

$$\varepsilon^2 = \frac{\frac{1}{L_1^{(e)} L_2^{(e)}} \sum \sum_{(i_1, i_2) \in \Theta^{(e)}} (u(i_1, i_2) - \hat{u}(i_1, i_2))^2}{\frac{1}{L_1^{(e)} L_2^{(e)}} \sum \sum_{(i_1, i_2) \in \Theta^{(e)}} u(i_1, i_2)^2 - \left( \frac{1}{L_1^{(e)} L_2^{(e)}} \sum \sum_{(i_1, i_2) \in \Theta^{(e)}} u(i_1, i_2) \right)^2}, \quad (4.40)$$

где кадр  $\Theta^{(e)}$  размером  $150 \times 150$  элементов расположен в центре кадра  $\Theta^{(s)}$  наблюдаемого изображения. Для рис. 4.29, *в* ошибка  $\varepsilon^2 = 11\%$ , а для рис. 4.29, *г* —  $\varepsilon^2 = 6\%$ . Таким образом, итерационный алгоритм практически полностью компенсирует краевые эффекты. Кроме того, применение экстраполяции позволяет снизить среднеквадратическую ошибку оценивания почти в 2 раза.

Наряду с описанными выше свойствами итерационные алгоритмы могут быть легко преобразованы в нелинейные путем введения нелинейных ограничений для восстанавливаемого изображения [4.6, 4.11]. Ограничения формулируются на основе априорных данных о форме или структуре объектов на исходном изображении. К априорным данным относятся такие свойства изображения, как неотрицательность яркости, ее верхний и нижний пределы, минимальная мощность сигнала, ограниченная пространственная и спектральная протяженность и т.п.

Даже учет такого простейшего ограничения, как верхний и нижний пределы значений яркости, приводит к значительному улучшению качества восстановления, поскольку среди всех возможных решений выбирается то, которое не имеет сильных осцилляций яркости.

Итерационный алгоритм, например (4.39), с ограничением имеет вид

$$\begin{aligned} \hat{u}^{(0)}(i_1, i_2) &= h_1(i_1, i_2) \otimes \otimes s(i_1, i_2); \\ \hat{u}^{(n)}(i_1, i_2) &= h_1(i_1, i_2) \otimes \otimes s(i_1, i_2) + (\delta(i_1, i_2) - h_2(i_1, i_2)) \otimes \otimes \mathfrak{Z} \{ \hat{u}^{(n-1)}(i_1, i_2) \}, \end{aligned} \quad (4.41)$$

где  $\mathfrak{Z}\{\cdot\}$  — оператор ограничения.

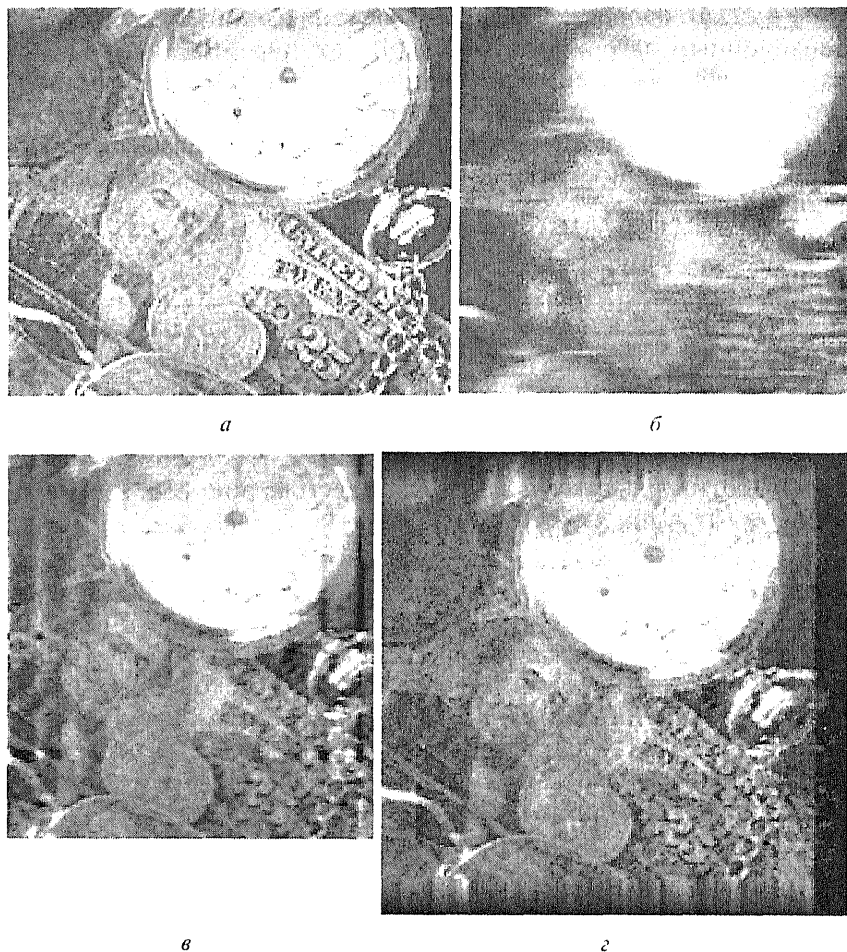


Рис. 4.29. Восстановление изображения «Часы» итерационным алгоритмом

Например, если используется оператор ограничения на неотрицательность  $\mathfrak{I}\{u(i_1, i_2)\} = \begin{cases} u(i_1, i_2); u(i_1, i_2) \geq 0, \\ 0; u(i_1, i_2) < 0, \end{cases}$  то из (4.41)

следует, что в тех областях, где яркость оценки  $\hat{u}^{(n-1)}(i_1, i_2)$  меньше нуля, изменение оценки не происходит. Нелинейный итерационный алгоритм (4.41) будет сходиться, если сходится линейный алгоритм (4.39) и оператор  $\mathfrak{I}\{\cdot\}$  является нерасши-

ряющим. Для пространства сигналов  $L_2$  это означает, что действие оператора на изображение не должно приводить к увеличению его энергии. Очевидно, что к нерасширяющимся операторам относятся оператор ограничения на неотрицательность, а также оператор ограничения диапазона, который определяется следующим соотношением:

$$\mathfrak{I}\{u(i_1, i_2)\} = \begin{cases} u(i_1, i_2); & a \leq u(i_1, i_2) \leq b, \\ a; & u(i_1, i_2) < a, \\ b; & u(i_1, i_2) > b. \end{cases} \quad (4.42)$$

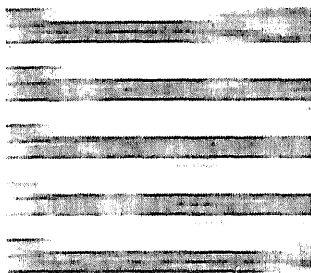
Для большинства цифровых изображений диапазон изменения яркости равен  $[0, 255]$ . Использование нелинейного алгоритма с ограничением диапазона для восстановления изображения «Часы» (рис. 4.29, б) обеспечивает уменьшение среднеквадратической ошибки до 4 % по сравнению с алгоритмом без ограничения. Особенно эффективен этот алгоритм при восстановлении изображений с распределением яркости, близким к бинарному.

На рис. 4.30 приведены результаты восстановления изображения «Текст», где  $a$  — исходное изображение размером  $98 \times 113$  элементов;  $b$  — часть исходного изображения, попадающая в кадр  $\Theta^{(s)}$  размером  $98 \times 98$  элементов;  $c$  — наблюдаемое изображение размером  $98 \times 98$  элементов, полученное в результате смаза исходного изображения ( $L_{i_2}^{(h)} = 15$ ,  $q^2 = 10^3$ );  $d$  — экстраполированное изображение размером  $128 \times 128$  элементов;  $e$  — изображения, восстановленные по экстраполированному изображению итерационным линейным алгоритмом ( $\varepsilon^2 = 11\%$ ) и итерационным алгоритмом с ограничением диапазона яркости ( $\varepsilon^2 = 4.5\%$ ). Ошибка вычислялась по кадру  $\Theta^{(r)}$  размером  $68 \times 68$  элементов. На рис. 4.30 размеры изображений увеличены в полтора раза.

Таким образом, использование априорных данных о диапазоне изменения яркости позволило уменьшить среднеквадратическую ошибку почти в 2.5 раза. Экстраполяция обеспечила восстановление изображения в пределах кадра исходного изображения, размеры которого больше размеров кадра  $\Theta^{(s)}$  наблюдаемого изображения.

Решение (4.20)  
Но запись итер  
Рассмотрим од  
Этот оператор  
Выражение (4

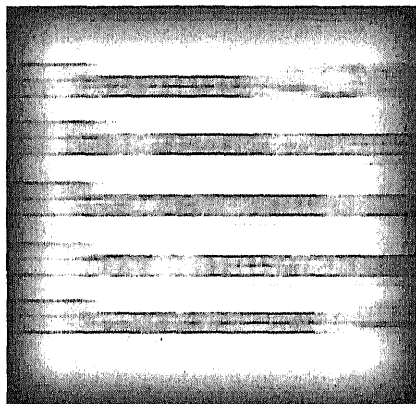
а



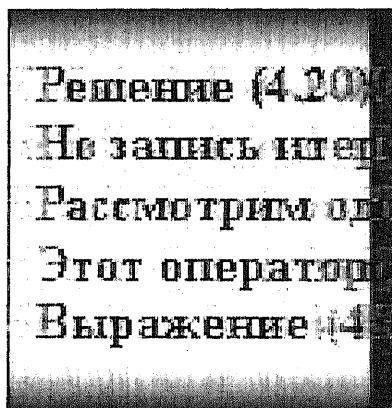
в

Решение (4.20)  
Но запись итер  
Рассмотрим од  
Этот оператор  
Выражение (4

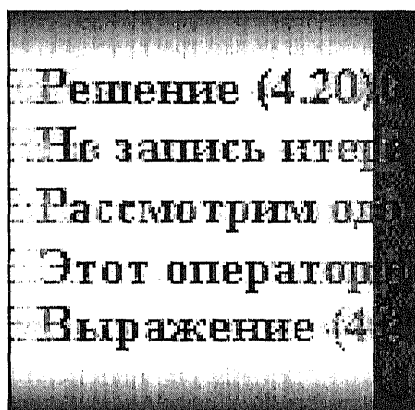
б



г



д



е

Рис. 4.30. Восстановление изображения «Текст» нелинейным итерационным алгоритмом

**ВОПРОСЫ К ГЛАВЕ 4**

4.1. Почему ухудшение четкости изображения описывается интегралом свертки исходного изображения и импульсной характеристикой искажающей системы?

4.2. В чем принципиальное различие между расфокусировкой и смазом?

4.3. Почему нельзя абсолютно точно восстановить изображение, сформированные реальными системами?

4.4. Чем отличается задача восстановления изображений алгебраическим методом при наличии шума наблюдения от аналогичной задачи при его отсутствии?

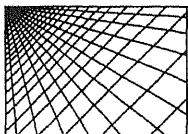
4.5. Назовите условия, при выполнении которых инверсная фильтрация обеспечивает высокое качество восстановления изображений.

4.6. Почему помехоустойчивость фильтра Винера выше, чем инверсного?

4.7. Назовите причины возникновения краевых эффектов и методы борьбы с ними.

4.8. Каковы достоинства и недостатки итерационных методов?

4.9. Почему итерационный алгоритм восстановления при гауссовской ФРТ отличается от итерационного алгоритма при цилиндрической ФРТ?



## ГЛАВА 5

### ГЕОМЕТРИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ И ПРИВЯЗКА ИЗОБРАЖЕНИЙ

Во многих задачах тематического дешифрирования применяется взаимное сопоставление между собой изображений, сформированных с помощью датчиков различных физических полей. Ярким примером может служить развитие дистанционных методов контроля природных ресурсов и динамики экосистем (так называемого мониторинга), т.е. сопоставления снимков одной и той же территории, полученных в разное время и/или с помощью различных датчиков. Чаще всего используются датчики, регистрирующие оптическое, радиолокационное, радиотепловое, магнитное и другие поля. Совместное использование различных физических полей требует предварительной обработки соответствующих им изображений, например, с целью перевода изображений в одну спектральную область.

На практике изображения одного и того же объекта или участка местности, полученные в разное время или с помощью различных датчиков, могут значительно различаться один от другого. Отсюда вытекает ряд важных задач привязки, а также точной взаимной геометрической и амплитудной коррекции для последующего совместного анализа. В любом случае это требует установления соответствия между элементами исходных изображений, что сводится к выделению так называемых *опорных* (по другому, *реперных* или *сопряженных*) точек на изображениях, по которым можно осуществить координатную привязку снимков с одновременной геометрической коррекцией. (Точки на двух изображениях называются *сопряженными*, если они являются образами одной точки сцены [5.1, гл.13].) Например, аэрокосмический компьютерный мониторинг предполагает наличие дискретного по времени наблюдения с небольшим временным интерва-



лом, и поэтому, когда движущаяся камера фиксирует яркостный образ наблюдаемого объекта (оптическую поверхность) в виде последовательности изображений, этот образ от снимка к снимку деформируется вследствие *перспективных* искажений и изменения положения камеры. Геометрия соответствующих деформаций моделируется *проективными преобразованиями*, которые составляют более обширный класс, нежели известные преобразования евклидовой геометрии (достаточно сказать, что длины и углы в проективной геометрии не сохраняются, а параллельные линии могут пересекаться! [5.12]).

Восстановление пространственного рельефа по стереоснимкам приводит к проблеме идентификации: установлению точного координатного (поточечного) соответствия элементов стереоизображений. Решение этой задачи состоит в выделении пар реперных фрагментов и оценивании параметров «расхождения» соответствующих точек (это именуется в стереофотограмметрии бинокулярной диспаратностью), по которым можно восстановить функцию геометрического преобразования и оценить поверхность трехмерной сцены (рельеф) [5.5].

### 5.1. ГЕОМЕТРИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ НА ПЛОСКОСТИ И В ПРОСТРАНСТВЕ

Геометрия является математическим базисом для решения многих задач машинного зрения и обработки изображений и содержит множество подобластей. Здесь мы рассмотрим лишь некоторые в охарактеризованном выше контексте привязки, преобразования и совмещения разновременных изображений одного и того же объекта.

При изучении геометрических преобразований плоских изображений (т.е. относящихся к двумерному случаю —  $2D$ ) будем предполагать, что мы работаем в евклидовом пространстве, где имеется ортонормированная декартова система координат, в которой координатные оси взаимно ортогональны, а соответствующие им единичные отрезки имеют одинаковую длину. Тогда каждой точке изображения ставится в соответствие упорядоченная пара чисел  $(x, y)$  декартовых координат: их можно интерпретировать как двумерный вектор  $x$ , геометрически представляемый отрезком прямой линии из точки  $(0, 0)$  в точку  $(x, y)$ .

Двумерные преобразования на плоскости будем интерпретировать как движения точек по отношению к фиксированному базису (а не как изменение базиса, оставляющее точки непод-

вижными). В частности, нас особенно будут интересовать *линейные преобразования*, представляемые матрицами, т. е. преобразования, при которых новые координаты точки линейно зависят от старых координат этой точки следующим образом:

$$\mathbf{x}' = \mathbf{T}\mathbf{x} . \quad (5.1)$$

Линейные преобразования могут быть различного типа, начиная от общего преобразования с произвольными элементами матрицы  $\mathbf{T}$  вплоть до специальных, случаев, когда на элементы матрицы накладываются те или иные ограничения. Интуитивно ясно, что каждому линейному преобразованию (или движению) на плоскости соответствует обратное, переводящее точки в первоначальное положение, и любым двум последовательно выполняемым преобразованиям точек плоскости соответствует некоторое третье преобразование, осуществляющее аналогичную (по результату) операцию. В таком случае принято говорить, что множество всех невырожденных линейных преобразований  $\mathbf{T}$  является замкнутым или, иначе, формирует группу, называемую здесь *общей линейной группой*. Интересно отметить, что само множество общих линейных преобразований может быть разбито на замкнутые подмножества или подгруппы. Прежде всего рассмотрим матрицы преобразования, связанные с наиболее важными подгруппами общей линейной (или *проективной*) группы, а именно, *евклидову* подгруппу, а также подгруппы *подобия* и *аффинную*. Это есть следствие того, что евклидова геометрия (так же как и аффинная) в действительности является подмножеством упомянутой нами выше проективной геометрии.

#### 5.1.1. ТОЧКИ И ПРЯМЫЕ ЛИНИИ НА ПЛОСКОСТИ – ДВОЙСТВЕННОСТЬ ОПИСАНИЙ

Прямая линия на плоскости, как известно из аналитической геометрии, состоит из всех точек, удовлетворяющих уравнению

$$ax + by + 1 = 0 .$$

Пусть две точки имеют координаты  $(x_1, y_1)$  и  $(x_2, y_2)$  соответственно. Каково уравнение линии, соединяющей их? Ясно, что, поскольку линия проходит через эти точки, она должна удовлетворять двум уравнениям:

$$ax_1 + by_1 + 1 = 0 ,$$

$$ax_2 + by_2 + 1 = 0 .$$

Данную систему из двух уравнений можно легко разрешить относительно неизвестных значений  $a$  и  $b$  и получить соответствующие выражения:

$$a = \frac{y_1 - y_2}{x_1 y_2 - x_2 y_1}, \quad b = \frac{x_2 - x_1}{x_1 y_2 - x_2 y_1}.$$

С другой стороны, предположим, что имеются две линии и нужно найти их точку пересечения  $(x, y)$ . Но две прямые должны соответствовать уравнениям

$$a_1 x + b_1 y + 1 = 0,$$

$$a_2 x + b_2 y + 1 = 0.$$

Отсюда для координат точки пересечения  $(x, y)$  получаем соотношения, аналогичные приведенным выше соотношениям для параметров линии  $(a, b)$ :

$$x = \frac{b_1 - b_2}{a_1 b_2 - a_2 b_1}, \quad y = \frac{a_2 - a_1}{a_1 b_2 - a_2 b_1}.$$

Здесь просматривается очень важная симметрия, или *двойственность* между проблемами пересечения двух прямых и (с другой стороны) линии, проходящей через две заданные точки. Координаты (параметры) пары линий и координаты пары точек в обоих случаях входят в формулы одинаковым образом. Далее мы увидим, что отмеченная двойственность распространяется и на другие соотношения между геометрическими объектами.

Имеется ряд проблем, связанных со специальными соотношениями выделенных пар точек и прямых. Предположим, что координаты двух точек отличаются лишь скалярным множителем:  $x_2 = \lambda x_1$ ,  $y_2 = \lambda y_1$ . Это означает, что  $x_1 y_2 - x_2 y_1 = 0$  и параметры прямой, соединяющей выделенные точки, определить невозможно. Прямая линия в данном случае проходит через начало координат  $(0, 0)$ , что, собственно, и создает проблему. Здесь нельзя непосредственно использовать уравнение прямой линии (проходящей через начало координат). Аналогичная проблема возникнет, когда мы попытаемся (формально, из приведенных выше уравнений) найти точку пересечения двух параллельных прямых, когда  $a_2 = \lambda a_1$ ,  $b_2 = \lambda b_1$ :

## 5.1.1.2. ОДНОРОДНЫЕ КООРДИНАТЫ

Для преодоления отмеченных проблем описания геометрических объектов, а также для решения задач преобразования 3D-пространства и 2D-плоскости в единообразном (матричном) виде вводится формализм так называемых *однородных координат*. Однородными координатами служат тройки чисел  $(\bar{x}, \bar{y}, w)$  (одновременно не равные нулю), связанные с обычными координатами точек плоскости соотношением:

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ w \end{pmatrix} = w \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \text{ так что } x = \frac{\bar{x}}{w}, \quad y = \frac{\bar{y}}{w}.$$

Совершенно очевидным свойством однородных координат является эквивалентность пары *однородных векторов*, если один в другой переводятся посредством скалярного множителя

$$(x, y, 1)' \equiv (\lambda x, \lambda y, \lambda)', \lambda \neq 0.$$

Поскольку скалярный множитель  $\lambda$  произвольный, однородные координаты в действительности представляют линию, проходящую через начало координат в евклидовом пространстве. Прямые линии на плоскости также можно представить трехмерными векторами в однородных координатах:

$$\begin{pmatrix} \bar{a} \\ \bar{b} \\ c \end{pmatrix} = c \begin{pmatrix} a \\ b \\ 1 \end{pmatrix}, \text{ где } c - \text{ произвольный скалярный множитель.}$$

Видно, что, как и для двух точек, однородные координаты двух линий эквивалентны, если отличаются лишь общим скалярным множителем. Однородные точки  $x = (\bar{x}, \bar{y}, w)'$ , лежащие на однородной линии  $p = (\bar{a}, \bar{b}, c)'$ , определяются уравнением

$$\bar{a}\bar{x} + \bar{b}\bar{y} + cw = 0, \text{ или } x^t p = 0.$$

Таким образом, точки и линии имеют здесь одинаковые представления. Нетрудно заметить, что прямым, проходящим через начало координат в данном представлении, соответствует значение  $c = 0$ . Двойственным образом точка пересечения двух параллельных прямых, лежащая в бесконечности, имеет множитель  $w = 0$ .

## 5.1.3. ЕВКЛИДОВЫ ПРЕОБРАЗОВАНИЯ

Сцену иногда можно рассматривать как твердое тело, когда взаимные деформации элементов сцены в трехмерном пространстве не допускаются. Аналогично и плоскость иногда можно считать жесткой (недеформируемой). Жестким движениям плоскости соответствует евклидова подгруппа, содержащая лишь преобразования сдвига и поворота (рис. 5.1), математически записываемыми в векторно-матричной форме как

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \mathbf{R} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}, \quad (5.2)$$

с матрицей поворота на угол  $\varphi$  вида  $\mathbf{R} = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$  и вектором трансляции (сдвига)  $\mathbf{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ .

При помощи троек однородных координат и матриц третьего порядка можно описать любое линейное преобразование плоскости. Действительно, введением дополнительного единичного компонента уравнение (5.2) можно переписать следующим образом:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^t & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (5.3)$$

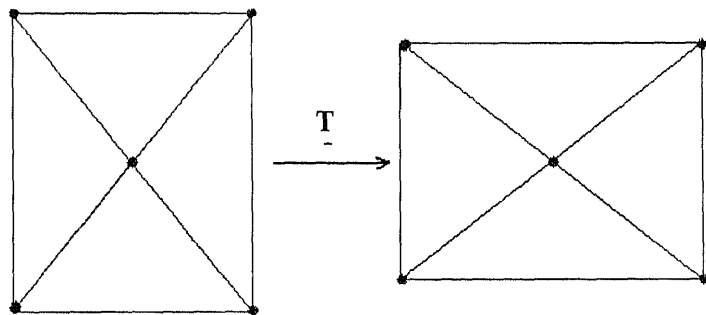


Рис. 5.1. Действие евклидова преобразования на пять точек плоскости (сдвиг, поворот)

Отметим далее, что два последовательно проведенных жестких движения плоскости могут быть представлены единственным движением:

$$\begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{t}_1 \\ \mathbf{0}' & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_2 & \mathbf{t}_2 \\ \mathbf{0}' & 1 \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{R}_1 \mathbf{R}_2 & \mathbf{R}_1 \mathbf{t}_2 + \mathbf{t}_1 \\ \mathbf{0}' & 1 \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}. \quad (5.4)$$

Комбинация двух последовательных вращений  $\mathbf{R}_1$  и  $\mathbf{R}_2$ , очевидно, сводится к вращению  $\mathbf{R}_1 \mathbf{R}_2$ . Кроме того, выбором вращения  $\mathbf{R}_2 = \mathbf{R}_1^{-1}$  и сдвига  $\mathbf{t}_2 = -\mathbf{R}_1^{-1} \mathbf{t}_1$  такое (второе) жесткое движение переводит точки плоскости в первоначальное положение. Отмеченной парой свойств, собственно говоря, и характеризуется группа, а класс матриц со структурой вида (5.3) известен как *евклидова группа* преобразований. (Она является, естественно, частным случаем линейных преобразований, у которых матрицы произвольные. Эти матрицы невырожденные и формируют общую линейную группу преобразований или *проективную группу*.) Интересно, что матрицы вращения  $\mathbf{R}$  сами по себе формируют так называемую *ортогональную* подгруппу с замечательным свойством  $\mathbf{R}^t \mathbf{R} = \mathbf{I}$ , где  $\mathbf{I}$  — единичная матрица.

#### 5.1.4. АФФИННЫЕ ПРЕОБРАЗОВАНИЯ

Если матрицу вращения в (5.2) заменить общей невырожденной матрицей  $\mathbf{A}$ , то получим преобразование

$$\begin{aligned} x' &= ax + by + c, \\ y' &= dx + ey + f, \end{aligned} \quad (5.5)$$

или в матричном виде

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{c}, \quad \mathbf{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a & b \\ d & e \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c \\ f \end{pmatrix},$$

и в однородных координатах

$$\begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{c} \\ \mathbf{0}^t & 1 \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}, \quad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (5.6)$$

Здесь также предполагается, что определитель матрицы преобразования не равен нулю:  $\det \mathbf{A} = \begin{vmatrix} a & b \\ d & e \end{vmatrix} = ae - bd \neq 0$ .

Уравнения (5.5), (5.6) определяют общую форму записи хорошо известного *аффинного преобразования* (рис. 5.2). Любое аффинное преобразование имеет *обратное*, которое также является аффинным. Произведение прямого и обратного преобразований дает *единичное преобразование*, оставляющее все на месте. Аффинное преобразование является самым общим взаимно однозначным отображением плоскости на плоскость, при котором сохраняются прямые линии. Сохраняются также отношения длин отрезков, лежащих на одной прямой (или на параллельных прямых), и отношения площадей фигур. Параллельные прямые сохраняют параллельность. В аффинных преобразованиях плоскости особую роль играют несколько важных частных случаев, имеющих простой и наглядный геометрический смысл, а также хорошо прослеживаемые геометрические характеристики [5.4, гл.15].

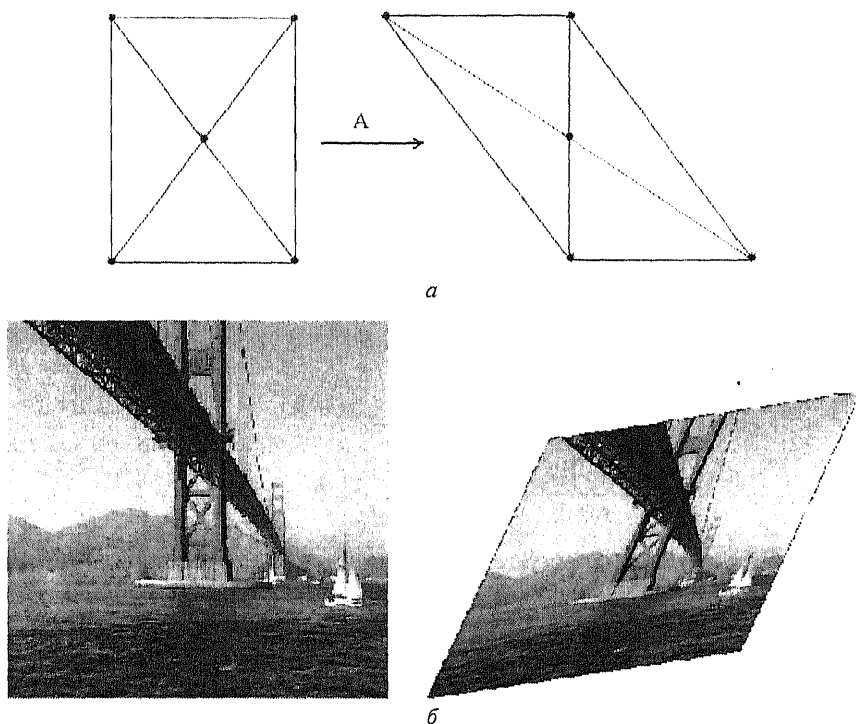


Рис. 5.2. Действие аффинного преобразования на пять точек (сдвиг, поворот, изменение масштабов вдоль осей, косоугольность с сохранением параллельных линий) (а); исходное изображение (слева) и его аффинно-преобразованная копия (справа); параметры аффинного преобразования:  $x' = 1.3x + 0.5y$ ;  $y' = 0.2x + 1.1y$  (б)

1. *Растяжение (сжатие)* вдоль координатных осей, задаваемое в виде

$$\begin{aligned}x' &= \alpha x, \quad \alpha > 0, \\y' &= \beta y, \quad \beta > 0.\end{aligned}$$

Растяжению вдоль определенной оси соответствует значение масштабного множителя, большего единицы. В однородных координатах матрица растяжения (сжатия) имеет вид

$$\mathbf{T}_{af} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

2. *Поворот* вокруг начальной точки на угол  $\varphi$ , описываемый формулой:

$$\begin{aligned}x' &= x \cos \varphi - y \sin \varphi, \\y' &= x \sin \varphi + y \cos \varphi.\end{aligned}$$

Матрица вращения (для однородных координат)

$$\mathbf{T}_{af} = \begin{bmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3. *Перенос*, задаваемый простейшими соотношениями:

$$\begin{aligned}x' &= x + c, \\y' &= y + f.\end{aligned}$$

Матрица переноса имеет вид

$$\mathbf{T}_{af} = \begin{bmatrix} 1 & 0 & c \\ 0 & 1 & f \\ 0 & 0 & 1 \end{bmatrix}.$$

4. *Отражение* (относительно какой-либо из осей, например оси абсцисс), задаваемое при помощи формулы:

$$\begin{aligned}x' &= x, \\y' &= -y.\end{aligned}$$

Матрица отражения, соответственно

$$\mathbf{T}_{af} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$



Из курса аналитической геометрии хорошо известно, что любое аффинное преобразование (5.5) всегда можно представить в виде композиции последовательно выполняемых простейших преобразований означенного вида. Более того, суперпозиция аффинных преобразований также является аффинным преобразованием. Ясно, что аффинные преобразования образуют *аффинную группу*. В частности, подгруппой аффинной группы преобразований является *группа подобия* (содержащая преобразования сдвига, поворота и изменения масштаба):

$$\begin{aligned}x' &= \alpha x \cos \varphi - \alpha y \sin \varphi + c, \\y' &= \alpha x \sin \varphi + \alpha y \cos \varphi + f.\end{aligned}$$

В то же время аффинная группа является подгруппой общей линейной (проективной) группы, а евклидова группа – частным случаем аффинной группы преобразований. Поэтому все отмеченные преобразования формируют иерархию в том смысле, что верно соотношение для их взаимной соподчиненности

$$\text{евклидово преобразование} \subset \text{аффинное} \subset \text{проективное преобразование}.$$

Зная параметры аффинного преобразования, можно вычислить непосредственно и параметры обратного преобразования:

$$\begin{aligned}x &= Ax' + By' + C, \\y &= Dx' + Ey' + F,\end{aligned}$$

решив систему уравнений (5.5) относительно  $(x', y')$ :

$$\begin{aligned}A &= \frac{e}{\det A}, \quad B = -\frac{b}{\det A}, \quad C = \frac{bf - ec}{\det A}, \\D &= -\frac{d}{\det A}, \quad E = \frac{a}{\det A}, \quad F = \frac{cd - af}{\det A}.\end{aligned}$$

Если параметры таковы, что  $(1 - a)(1 - e) \neq bd$ , то данное аффинное преобразование имеет *неподвижную точку*  $(X_0, Y_0)$ :

$$\begin{aligned}X_0 &= \frac{bf + c(1 - e)}{(1 - a)(1 - e) - bd}, \\Y_0 &= \frac{cd + f(1 - a)}{(1 - a)(1 - e) - bd}.\end{aligned}$$

Заметим, что при  $x' = Ax$ ,  $c = 0$  начало координат и будет являться неподвижной точкой.

## 5.1.5. ПРОЕКТИВНЫЕ ПРЕОБРАЗОВАНИЯ

Как выше уже было сказано, общими линейными преобразованиями  $T$  (в представлении однородными координатами)

$$\begin{pmatrix} \bar{x}' \\ \bar{y}' \\ w' \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{y} \\ w \end{pmatrix} \quad (5.7)$$

формируется группа проективных преобразований (рис. 5.3). При представлении в обычных координатах очевидно, что соотношение (5.7) будет иметь нелинейный вид, связанный с перенормировкой:

$$x' = \frac{t_{11}x + t_{12}y + t_{13}}{t_{31}x + t_{32}y + t_{33}}, y' = \frac{t_{21}x + t_{22}y + t_{23}}{t_{31}x + t_{32}y + t_{33}}.$$

Проективные преобразования в общем случае не сохраняют параллельности линий. Свойством, сохраняющимся при проективном преобразовании, является так называемая *коллинеарность* точек: три точки, лежащие на одной прямой (т.е. коллинеарные), после преобразования остаются лежать на одной прямой (рис. 5.3). Поэтому обратимое проективное преобразование принято называть еще *коллинеацией* (гомографией).

Проективное преобразование связано с отображением трехмерного изображения на двумерную плоскость. С математической точки зрения, удобно рассматривать мир, включенный в трехмерное проективное пространство  $P^3$ , а плоскость изображения, включенную в проективное двумерное пространство, —  $P^2$ . Точки на трехмерной сцене и на изображении представляются в проективных пространствах как векторы в однородных координатах.

Проективное преобразование из  $P^3$  в  $P^2$  (перспективная проекция), отображающее евклидову точку сцены  $\mathbf{p} = (X, Y, Z)^t$  в точку изображения  $\mathbf{x} = (x, y)^t$  и выраженное в однородных координатах, задается в виде

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ w \end{pmatrix} = \begin{pmatrix} wx \\ wy \\ w \end{pmatrix} = \mathbf{P} \begin{pmatrix} WX \\ WY \\ WZ \\ W \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \\ \bar{W} \end{pmatrix}. \quad (5.8)$$

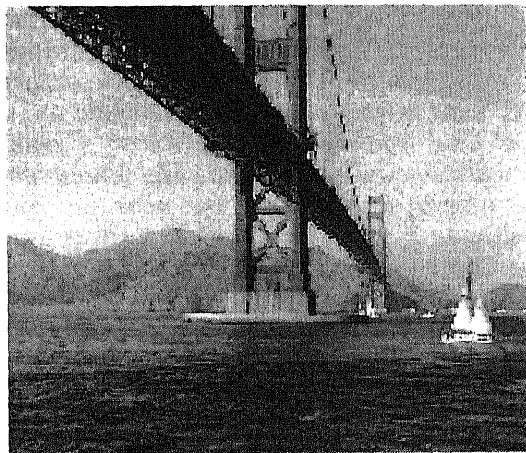
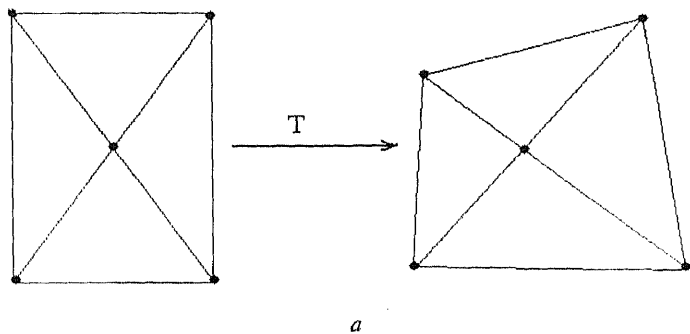
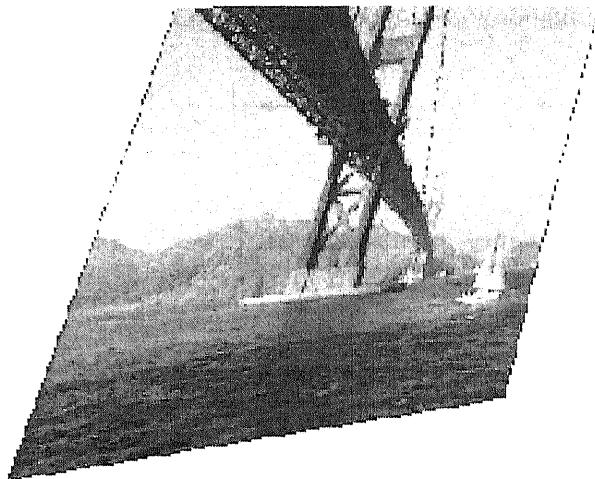


Рис. 5.3. Действие проективного преобразования на пять точек плоскости (a); исходное изображение (б); проективно преобразованные образы (в) и (г); параметры проективного преобразования, соответствующие (в) и (г):

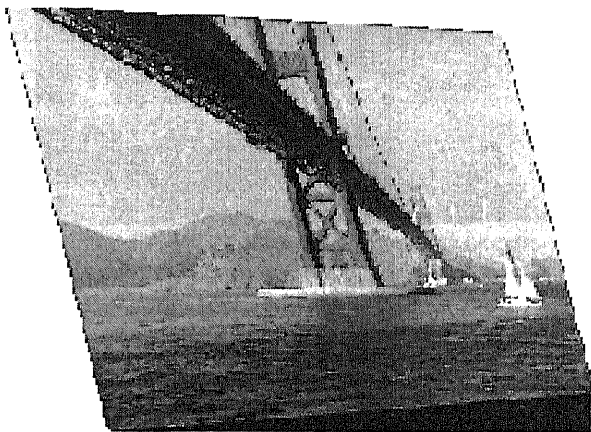
$$x' = \frac{1.1x + 0.35y}{0.00075x + 0.0005y + 1}, y' = \frac{0.2x + 1.1y}{0.00075x + 0.0005y + 1} \text{ и}$$

$$x' = \frac{1.1x - 0.2y}{0.00075x + 0.00005y + 1}, y' = \frac{0.1x + 0.9y}{0.00075x + 0.00005y + 1}$$

(см. также с. 142)



в



г

*Рис. 5.3. Окончание*

Однородные координаты векторов проективного пространства  $(\bar{X}, \bar{Y}, \bar{Z}, \bar{W})'$  и проективной плоскости  $(\bar{x}, \bar{y}, w)'$  соотносятся с неоднородными (евклидовыми) координатами векторов  $\mathbf{p}$  и  $\mathbf{x}$  следующим образом:

$$\mathbf{p} = \left( \frac{\bar{X}}{\bar{W}}, \frac{\bar{Y}}{\bar{W}}, \frac{\bar{Z}}{\bar{W}} \right)' \quad \text{и} \quad \mathbf{x} = \left( \frac{\bar{x}}{w}, \frac{\bar{y}}{w} \right)'.$$

Проективная геометрия составляет математический базис машинного зрения и компьютерной графики. Основные области применения связаны с описанием как процесса формирования изображений, так и их инвариантного представления, а именно: калибровка регистрирующей камеры, анализ движения по серии изображений, распознавание образов, реконструкция сцен по стереоснимкам, синтез изображений, анализ и восстановление формы по полутонам. Полезно отметить тот факт, что композиция двух перспективных проекций не обязательно является перспективной проекцией, но определяет проективное преобразование; т.е. (как мы знаем) проективные преобразования формируют группу, в то время как перспективные проекции — нет.

В связи с этим напомним, что изображение объектов на снимке, сформированном регистрирующей камерой, связано с чрезвычайно важной геометрической операцией — проектированием при помощи пучка прямых, поскольку каждая 2D точка является проекцией множества 3D точек вдоль некоторого направления («луча проектирования») в плоскость снимка (рис. 5.4). Предположим, что плоскость снимка камеры в системе координат  $(X, Y, Z)$  определяется соотношением  $Z = f$ . Если расстояние от плоскости изображения до центра проекции равно  $f$ , то координаты элементов изображения  $(x_i, y_i)$  соотносятся с пространственными координатами  $(X_i, Y_i, Z_i)$  объекта следующим образом:

$$x_i = \frac{fX_i}{Z_i}, \quad y_i = \frac{fY_i}{Z_i}. \quad (5.9)$$

Это нелинейные уравнения. Они могут быть сделаны линейными введением однородных координат. Заметим, что луч, проходящий через 2D точку  $(x_i, y_i)$ , является направляющим вектором

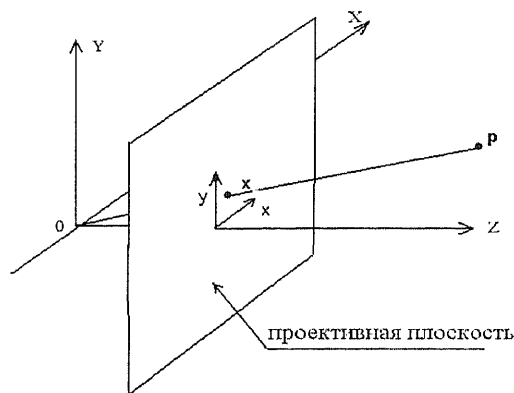


Рис. 5.4. Перспективная проекция

прямой, соединяющей точки  $(0, 0, 0)$  и  $(x_i, y_i, f)$ . В то же время 3D точка  $\lambda(x_i, y_i, f) = (\lambda x_i, \lambda y_i, \lambda f) = (X_i, Y_i, Z_i)$  также лежит на этом луче (и представляет его). При  $Z_i$  произвольном (пробегающем всевозможные значения) получим координаты всех 3D точек  $(\lambda x_i, \lambda y_i, Z_i) = (X_i, Y_i, Z_i)$  на этом луче, соответствующих единственной точке  $(x_i, y_i) = \left( \frac{fX_i}{Z_i}, \frac{fY_i}{Z_i} \right)$  на изображении в плоскости

$Z = f$ . По существу, каждая точка на изображении определяет луч, идущий от сцены в начало координат. Следовательно, однородные координаты точки (на снимке) в действительности представляют (и определяют) линию, проходящую через начало в евклидовом трехмерном пространстве  $E^3$ . Такой набор всевозможных линий, проходящих через начало  $E^3$ , и формирует *проективное* 2D-пространство  $P^2$  в плоскости  $Z = f$ . (Можно, конечно, положить фокусное расстояние  $f = 1$ , поскольку различным значениям  $f$  соответствует разный масштаб изображения.) В однородных координатах уравнения (5.9), естественно, имеют вид (5.8)

$$\begin{pmatrix} wx_i \\ wy_i \\ w \end{pmatrix} = \begin{pmatrix} \bar{x}_i \\ \bar{y}_i \\ \bar{w} \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \quad (5.9')$$

**Замечание.** При  $w = 0$ ,  $3D$  точка  $(\bar{x}, \bar{y}, 0)$  определяет линию, параллельную плоскости  $Z = 1$  (т.е. не имеющую с ней точек пересечения) и, следовательно, не имеет соответствия с какой-либо конечной точкой изображения. Такие линии, или однородные векторы, могут тем не менее иметь смысл, если считать, что соответствующая им точка  $(x, y)$  удаляется на бесконечность в направлении, задаваемом этими координатами:

$$\lim_{w \rightarrow 0} \left( \frac{\bar{x}}{w}, \frac{\bar{y}}{w}, 1 \right) = (x, y, 1) \Rightarrow \lim_{w \rightarrow 0} (\bar{x}, \bar{y}, w) = (\bar{x}, \bar{y}, 0).$$

Мы можем добавить все такие точки к проективной плоскости. Эти точки называются «идеальными» или точками на бесконечности. На изображениях проективной плоскости добавленные точки на бесконечности формируют «линию горизонта» (рис. 5.5). Существует разделение идеальных точек, обусловленное различными направлениями на плоскости; например, точки  $(1, 0, 0)$  и  $(0, 1, 0)$  связаны с горизонтальным и вертикальным направлениями (осями координат) соответственно. Можно также сказать, что все идеальные точки лежат на линии, называемой «идеальной линией» или *линией на бесконечности*, которая рассматривается тем не менее, как и обычная линия. Идеальная линия представляется в виде  $(0, 0, 1)$ .

**Определение.** Проективная плоскость  $P^2$  является аффинной плоскостью с присоединенными идеальной линией и множеством

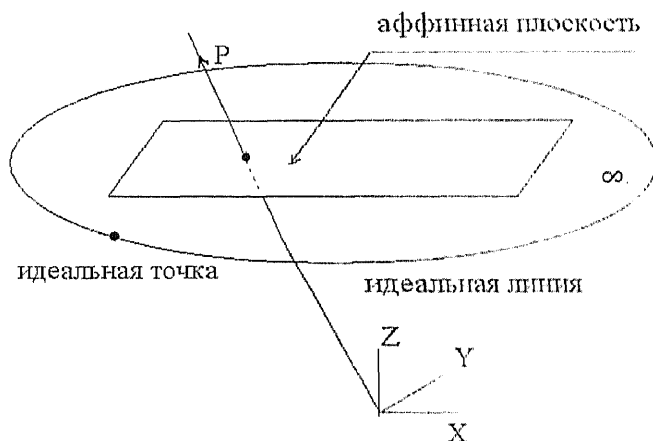


Рис. 5.5. Проективная плоскость = аффинная плоскость +  
+ идеальные точки (идеальная линия)

идеальных точек, которые не отличаются от обычных линий и точек [5.12]. (Аффинная плоскость, естественно, состоит из тех же точек, что и евклидова плоскость. Различие заключается в том, что в первой допускаются неоднородное масштабирование и косоугольность.)

**Замечание.** Найдём пересечение двух прямых  $\mathbf{u}_1 = (a_1, b_1, c_1)$  и  $\mathbf{u}_2 = (a_2, b_2, c_2)$ . Несложные вычисления показывают, что однородные координаты точки пересечения  $\mathbf{p}$  равны  $(b_1c_2 - b_2c_1, a_2c_1 - a_1c_2, a_1b_2 - a_2b_1)$ . Последняя формула легко запоминается, поскольку есть не что иное, как векторное произведение:  $\mathbf{p} = \mathbf{u}_1 \times \mathbf{u}_2$ . Если эти две прямые линии параллельны, т.е.  $\frac{a_1}{b_1} = \frac{a_2}{b_2}$ , то точка пересечения существует (идеальная точка!) и равна  $(b_1c_2 - b_2c_1, a_2c_1 - a_1c_2, 0)$ . Двойственным образом, задавая две точки  $\mathbf{p}_1$  и  $\mathbf{p}_2$ , можно непосредственно найти прямую, проходящую через них:  $\mathbf{u} = \mathbf{p}_1 \times \mathbf{p}_2$ .

Таким образом, идеализация процесса формирования изображения камерой может быть представлена как перспективная проекция из  $P^3$  в  $P^2$ . Допустим, что 3D координаты точек объекта известны. Тогда, зная элементы матрицы  $\mathbf{P}$ , относящиеся к данному проективному преобразованию, точки пространственного объекта можно связать с соответствующими им координатами на снимке в виде (5.8):

$$\begin{pmatrix} \lambda x_i \\ \lambda y_i \\ \lambda \end{pmatrix} = \mathbf{P} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}. \quad (5.10)$$

Очевидно, что неизвестный масштабный множитель определяется как

$$\lambda = p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34},$$

так что координаты изображения объекта имеют вид отношения

$$\begin{aligned} x_i &= \frac{p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}, \\ y_i &= \frac{p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}. \end{aligned} \quad (5.11)$$



Пусть координаты характерных элементов изображения объекта  $(x_i, y_i), i = 1, 2, \dots, N$  соотнесены с пространственными координатами точек объекта и требуется вычислить элементы матрицы  $\mathbf{P}$ , т.е. осуществить так называемую *калибровку* камеры (см. главу 6). По крайней мере, 6 точек объекта нужно идентифицировать для этого на снимке ( $N = 6$ ), поскольку элементы матрицы  $\mathbf{P}$  определены с точностью до масштабного множителя (только лишь отношения элементов  $p_{ij}$  значимы) и существует 11 степеней свободы (неизвестных параметров). Обычно же требуется большее число точек, так как измерения сопровождаются помехами и оптимальное решение, минимизирующее их влияние на результат, находится методом наименьших квадратов.

*Проективным базисом* на плоскости  $P^2$  является множество из 4 точек, таких, что любые три из них не лежат на прямой. Соответственно  $3 \times 3$  матрица, сформированная векторами-столбцами однородных координат любых трех точек, должна иметь полный ранг. Легко проверить, например, что точки  $\{(1, 0, 0)', (0, 1, 0)', (0, 0, 1)', (1, 1, 1)'\}$  формируют так называемый *канонический базис*. В соответствии с приведенным выше *замечанием*, канонический базис содержит точки на бесконечности вдоль каждой координатной оси  $\{(1, 0, 0)', (0, 1, 0)'\}$ , начало  $(0, 0, 1)'$  и единичную точку  $(1, 1, 1)'$ . Очевидно, что коллинеация на  $P^2$ :  $\mathbf{x}' = \mathbf{T}\mathbf{x}$  целиком характеризуется геометрически ее действием на точки базиса.

Элементы  $3 \times 3$  матрицы  $\mathbf{T}$  определены с точностью до масштабного множителя, и из них лишь восемь значений независимы. Поэтому, поскольку каждая точка содержит две независимые координаты, четыре пары сопряженных точек (на двух снимках) позволяют определить  $\mathbf{T}$ .

Мы знаем, что аффинная плоскость, в отличие от проективной плоскости, не содержит идеальных точек. Тогда точка  $(x, y, 0)'$  должна быть трансформируема в точку  $(\lambda x, \lambda y, 0)'$  для произвольного масштабного множителя  $\lambda$ :

$$\lambda \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \mathbf{T} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix},$$

что влечет  $t_{31} = t_{32} = 0$ . Матрица аффинного преобразования имеет вид

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ 0 & 0 & t_{33} \end{pmatrix},$$

и здесь содержится лишь шесть независимых параметров, поскольку масштаб также не важен.

### 5.1.6. ПОЛИНОМИАЛЬНОЕ ПРЕОБРАЗОВАНИЕ

Выше мы рассмотрели геометрические преобразования, в общем-то, идеализированной модели камеры. В действительности формирование изображений сопровождается различного рода нелинейными искажениями (типа оптической дисторсии линзы). Приведение текущих снимков друг к другу или к некоторому эталонному в таком случае можно осуществить лишь нелинейной функцией преобразования. Кроме того, неравномерность движения носителя регистрирующей камеры также приводит к тому, что на практике геометрические искажения снимков не устраняются аффинным преобразованием координат элементов снимка. Поэтому используют полиномиальную аппроксимирующую функцию преобразования (рис. 5.6)

$$\begin{aligned} x' &= a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 + \dots + a_p y^m, \\ y' &= b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + \dots + b_p y^m, \end{aligned} \quad (5.12)$$

где  $(x, y)$  — координаты точек эталонного снимка,  $(x', y')$  — соответствующие им координаты на текущем (сопоставляемом) снимке.

### 5.1.7. ОЦЕНИВАНИЕ ПАРАМЕТРОВ ПРЕОБРАЗОВАНИЯ

Параметры линейных и нелинейных преобразований (5.2), (5.5), (5.12) устанавливаются по парам взаимно соответствующих реперных точек, идентифицируемых в процессе поиска (см. далее). После этого каждой точке  $(x, y)$  эталонного снимка ставится в соответствие точка  $(x', y')$  текущего снимка. Будем считать, что точки цифрового изображения (пиксели) представлены на дискретной  $(M \times M)$  решетке с постоянным шагом так, что их целочисленные координаты имеют вид  $x_i = i - (M+1)/2$ ,  $y_j = j - (M+1)/2$ ;  $i, j = 1, 2, \dots, M$ . Коэффициенты  $(a_k, b_k)$  выбирают таким образом, чтобы минимизировать среднеквадратическую ошибку аппрок-

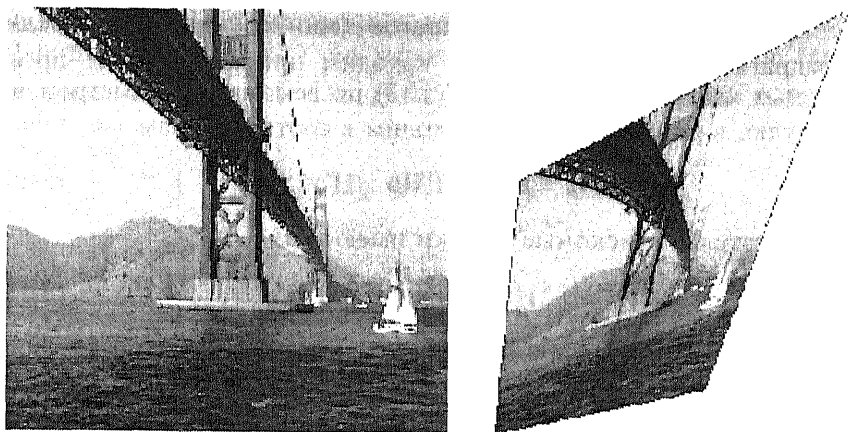


Рис. 5.6. Действие билинейного геометрического преобразования на исходный снимок; параметры преобразования:  
 $x' = 0.1x + 0.9y + 0.002xy$ ,  $y' = 0.2x + 1.1y + 0.0022xy$

симации фактически «наблюдаемых» координат  $(x'_i, y'_j)$  их полиномиальной оценкой  $(\hat{x}'_i, \hat{y}'_j)$  из (5.12) для набора заданных узловых точек  $(x_i, y_j)$ ,  $i, j \in (1, 2, \dots, N)$ . Координаты в плоскости наблюдаемого (текущего, или контролируемого) изображения удобно выразить в виде векторов

$$\mathbf{x}' = (x'_1, x'_2, \dots, x'_N)^t, \quad \mathbf{y}' = (y'_1, y'_2, \dots, y'_N)^t.$$

Аналогично коэффициенты полиномов можно представить в векторной форме:

$$\mathbf{b} = (b_1, b_2, \dots, b_p)^t,$$

$$\mathbf{a} = (a_1, a_2, \dots, a_p)^t.$$

Среднеквадратическую ошибку оценивания можно записать в компактной матричной форме

$$E^2 = (\mathbf{x}' - \mathbf{H}\mathbf{a})^t(\mathbf{x}' - \mathbf{H}\mathbf{a}) + (\mathbf{y}' - \mathbf{H}\mathbf{b})^t(\mathbf{y}' - \mathbf{H}\mathbf{b}), \quad (5.13)$$

где  $\mathbf{H} = \begin{pmatrix} 1 & x_1 & y_1 & \dots & y_1^m \\ 1 & x_2 & y_2 & \dots & y_2^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & y_N & \dots & y_N^m \end{pmatrix}$  — регрессионная матрица, соответствующая (5.12).

Оценки наименьших квадратов (минимизирующих средне-квадратическую ошибку  $E^2$ ) находим приравниванием производных квадратичной формы (5.13) по векторам параметров  $\mathbf{a}$  и  $\mathbf{b}$  нулю, в результате чего приходим к соотношениям

$$\mathbf{H}^t \mathbf{H} \mathbf{a} = \mathbf{H}^t \mathbf{x}', \quad \mathbf{H}^t \mathbf{H} \mathbf{b} = \mathbf{H}^t \mathbf{y}'. \quad (5.14)$$

Следовательно, искомые оценки имеют вид

$$\hat{\mathbf{a}} = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{x}', \quad \hat{\mathbf{b}} = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{y}'. \quad (5.15)$$

При больших значениях  $m$  регрессионная матрица  $\mathbf{H}$ , соответствующая (5.12), становится неустойчивой, что приводит к большим ошибкам в определении коэффициентов преобразования. Одним из способов уменьшения этого эффекта является использование полиномов Чебышева. Полиномиальное преобразование (5.12) в этом случае представляется как

$$\widehat{\mathbf{x}}' = \sum_{l=0}^m \sum_{k=0}^{m-l} \alpha_{lk} P_N^{(k)}(x) P_N^{(l)}(y), \quad (5.16)$$

$$\widehat{\mathbf{y}}' = \sum_{l=0}^m \sum_{k=0}^{m-l} \beta_{lk} P_N^{(k)}(x) P_N^{(l)}(y), \quad (5.17)$$

где  $P_N^{(l)}(x)$  — попарно ортогональные на заданном множестве точек многочлены Чебышева, получаемые из последовательности  $1, x, x^2, x^3, \dots$  методом ортогонализации Грама-Шмидта [5.3, гл.20]. Ортогональные многочлены до третьей степени имеют вид

$$P_N^{(0)}(x) = 1, P_N^{(1)}(x) = \lambda_N^{(1)} x, P_N^{(2)}(x) = \lambda_N^{(2)} [x^2 - (N^2 - 1)/12],$$

$$P_N^{(3)}(x) = \lambda_N^{(3)} [x^3 - (3N^2 - 7)x/20],$$

$$\lambda_N^{(1)} = [12/(N^2 - 1)]^{1/2}, \lambda_N^{(2)} = [4 \cdot 9 \cdot 5 / (N^2 - 1)(N^2 - 4)]^{1/2},$$

$$\lambda_N^{(3)} = [16 \cdot 25 \cdot 7 / (N^2 - 1)(N^2 - 4)(N^2 - 9)]^{1/2}.$$

Коэффициенты  $(\alpha_{kl}, \beta_{kl})$  снова выбирают таким образом, чтобы минимизировать среднеквадратическую ошибку аппроксимации фактически «наблюдаемых» координат  $(x'_i, y'_j)$  их полиномиальной оценкой  $(\widehat{x}'_i, \widehat{y}'_j)$  из (5.16), (5.17) для набора заданных узловых точек  $(x_i, y_j)$ ,  $i, j \in (1, 2, \dots, N)$ .

Пользуясь этой методикой, можно без труда вычислить коэффициенты  $(\alpha_{kl}, \beta_{kl})$  по методу наименьших квадратов [аналогично формулам (5.13)...(5.15)], где регрессионная матрица имеет вид

$$H' = \begin{pmatrix} 1 & P_N^{(1)}(x_1) & P_N^{(1)}(y_1) & \dots & P_N^{(m)}(y_1) \\ 1 & P_N^{(1)}(x_2) & P_N^{(1)}(y_2) & \dots & P_N^{(m)}(y_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & P_N^{(1)}(x_N) & P_N^{(1)}(y_N) & \dots & P_N^{(m)}(y_N) \end{pmatrix}. \quad (5.18)$$

## 5.2. ВОССТАНОВЛЕНИЕ ИЗОБРАЖЕНИЯ В ПРЕОБРАЗОВАННЫХ КООРДИНАТАХ

После оценивания параметров геометрического преобразования встает задача собственно геометрической коррекции или, по другому, восстановления изображения в преобразованных координатах.

Будем считать, что заданы два снимка ( $U$  и  $V$ ) одной и той же местности, полученные с некоторыми отклонениями точек съемки и условий освещенности. Вследствие этого изображения на снимках отличаются друг от друга геометрическими искажениями. Будем также считать, что на изображениях выделены сопряженные точки, по которым произведено оценивание параметров геометрического преобразования. Зная коэффициенты линейного (или полиномиального) преобразования, можно вычислить в плоскости корректируемого изображения  $V$  координаты всех точек  $(x'_i, y'_j)$ , соответствующие точкам с целочисленными координатами  $(x_i, y_j)$  на эталонном снимке  $U$  (рис. 5.7).

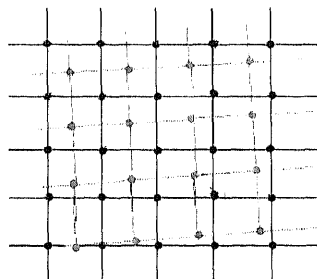


Рис. 5.7. Вычисленные координаты  $(x', y')$  (выделены серым тоном), наложенные на исходную дискретную целочисленную решетку

Восстановив уровни яркости наблюдаемых элементов в вычисленных точках на корректируемом снимке, т. е. осуществив «передискретизацию», полученные значения также можно поместить на дискретном растре размером  $N \times N$ , приведя тем самым искаженное изображение  $V$  в формат эталонного снимка  $U$ .

Поскольку координаты  $(x'_i, y'_j)$  не попадают чаще всего в узлы дискретной решетки (рис. 5.7), возникает задача восстановления соответствующего значения яркости по ближайшим отсчетам. Она решается с помощью методов двумерной интерполяции [5.2, разд. 5.3]. Интерполированное непрерывное изображение в плоскости снимка  $V$  можно описать сверткой

$$\hat{v}(x, y) = \sum_i \sum_j v(i, j) g(x - ih, y - jh), \quad (5.19)$$

где  $g(t, s)$  — интерполирующая функция (называемая также интерполяционным ядром);  $h$  — шаг дискретизации исходного изображения;  $v(ih, jh)$  — известные отсчеты яркости в точках дискретного раstra. Оценка непрерывного изображения позволяет осуществить его передискретизацию на новом множестве точек.

Интерполяционное ядро имеет значительное влияние на численное поведение интерполированных функций. Теоретически оптимальную интерполяцию обеспечивает известная sinc-функция, в одномерном случае имеющая вид

$$\sin c(x - ih) = \frac{\sin 2\pi\Omega(x - ih)}{2\pi(x - ih)}, \quad (5.20)$$

где  $\Omega$  есть ширина полосы частот  $v(x, y)$ . Из теоремы отсчетов следует (см. главу I), что sinc-функция дает наилучшую реконструкцию  $v(x, y)$ , если последняя имела ограниченный спектр и была первоначально оцифрована вблизи частоты Найквиста.

Поскольку интерполяция противоположна дискретизации, интерполирующая функция (5.20) по существу является идеальным низкочастотным фильтром, вырезающим основной участок ограниченного спектра из множества его повторяющихся копий. Однако этот теоретический метод практически невозможно реализовать в контексте обработки изображений. В частности, ограничение области суммирования в (5.19) приводит к тому, что осцилляции, известные как феномен Гиббса, будут проникать в восстанавливаемый образ  $\hat{v}(x, y)$ . Поэтому на практике используют интерполяционные ядра, реализация которых сопряжена с меньшими трудностями. В одномерном случае это прямоугольные, треугольные, B-сплайн функции и т.п. [5.2, разд. 4.3]. При выборе соответствующего ядра исходят из соображений как необходимой точности интерполяции, так и вычислительной эф-

фективности. Понятно, что здесь одномерные функции должны быть преобразованы в двумерные функции. Общий подход состоит во введении так называемых «сепарабельных» интерполяционных функций в виде произведения двух одномерных функций. Сепарабельность во многих отношениях достаточно привлекательна в приложениях, хотя и влечет неизотропность (за исключением гауссовых функций). Однако данные на квадратной решетке дискретизованы также неизотропно.

С вычислительной точки зрения предпочтителен алгоритм, известный как интерполятор по ближайшему соседу, где значение в точке  $(x, y)$  приравнивается значению ближайшего отсчета дискретного раstra. Этот метод соответствует прямоугольному интерполирующему ядру (рис. 5.8). Свертка с прямоугольной функцией в пространственной области эквивалентна умножению сигнала в области частот на sinc-функцию. Последняя является плохим приближением к низкочастотному фильтру, поскольку имеет бесконечное множество боковых лепестков. Алгоритм ближайшего соседа приводит к локальным сдвигам относительно первоначального изображения на величины разностей между вычисленной точкой и ближайшей точкой дискретного раstra (т.е. вплоть до  $\frac{h}{2}$ ). Треугольное ядро (рис. 5.9) в двумерном случае приводит к билинейной интерполяции по четырем ближайшим соседям точки  $(x, y)$ :  $\{[ih, jh], [(i+1)h, jh], [ih, (j+1)h], [(i+1)h, (j+1)h]\}$ .

Справа — график модуля фурье-образа ядра. (Пунктирной линией показан идеальный низкочастотный фильтр с частотой среза  $(-\pi, +\pi)$ .)

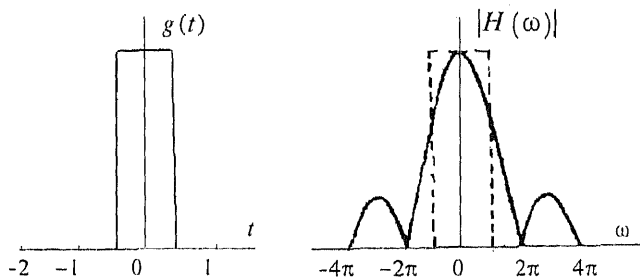


Рис. 5.8. Интерполятор по ближайшему соседу с прямоугольным ядром

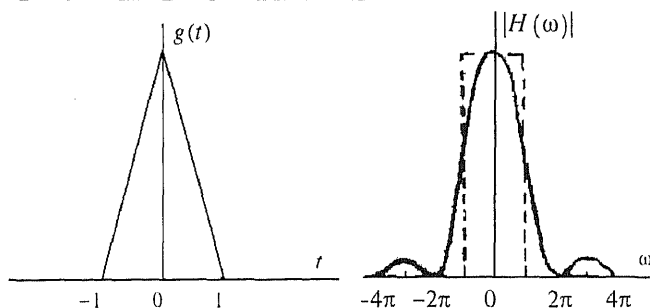


Рис. 5.9. Линейная интерполяционная функция и модуль ее фурье-образа (на правом рисунке пунктиром отмечен идеальный низкочастотный фильтр)

Здесь интерполированный сигнал представляется в виде

$$\hat{v}(x, y) = a + b(x - ih) + c(y - jh) + d(x - ih)(y - jh), \quad (5.21)$$

где

$$\begin{aligned} a &= v(i, j), b = v((i+1), j) - a, c = v(i, (j+1)) - a, \\ d &= v((i+1), (j+1)) + a - b - c. \end{aligned}$$

Приближение к низкочастотному фильтру в данном случае еще далеко от идеального, и к тому же производная интерполированного сигнала терпит разрывы в узлах интерполяции (тем не менее формула (5.21) часто применяется на практике, поскольку удовлетворяет одновременно требованиям приемлемой точности и приемлемым затратам вычислительных ресурсов).

Наиболее подходящим для интерполяции изображений является кубический *B*-сплайн (рис. 5.10), поскольку в результате его применения получается функция, непрерывная и гладкая в узлах интерполяции. Ядро кубической свертки составляется из кусков кубических полиномов, определенных на подынтервалах  $(-2, -1)$ ,  $(-1, 0)$ ,  $(0, 1)$ ,  $(1, 2)$  по каждой из координат. Вне интервала  $(-2, 2)$  интерполяционное ядро равно нулю. Двумерный кубический *B*-сплайн может быть записан как произведение двух одномерных интерполяционных функций по каждой из координат

$$g(s, t) = g(s)g(t),$$

где  $(s, t) = \left( \frac{x}{h} - i, \frac{y}{h} - j \right)$ , так что



$$g(s) = \begin{cases} \frac{|s^3|}{2} - s^2 + \frac{2}{3}, & |s| \in (0, 1); \\ -\frac{|s^3|}{6} + s^2 - 2|s| + \frac{4}{3}, & |s| \in (1, 2), \end{cases} \quad (5.22)$$

и  $g(s) = g(-s)$ , т.е. ядро симметричное.

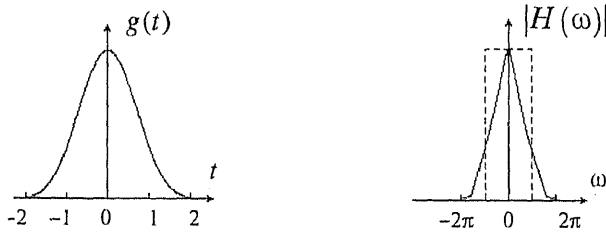


Рис. 5.10. Кубический В-сплайн и модуль его фурье-образа

Как показали непосредственные исследования, кубический В-сплайн имеет тенденцию к сглаживанию передискретизованного изображения по сравнению с его первоначальной копией. Поэтому были предприняты определенные усилия для выбора кубического сплайна, более подходящего задачам обработки изображений. Общий кубический сплайн задается в виде

$$\begin{aligned} g(s) &= a_{30} |s^3| + a_{20} s^2 + a_{10} |s| + a_{00}, & |s| \in (0, 1); \\ g(s) &= a_{31} |s^3| + a_{21} s^2 + a_{11} |s| + a_{01}, & |s| \in (1, 2). \end{aligned} \quad (5.23)$$

Есть несколько естественных ограничений на данное интерполяционное ядро. Так требуется, чтобы значение интерполирующей функции в нуле было равно 1, а в точках с координатами 1 и 2 равно 0. Кроме того, необходимо, чтобы ядро было непрерывным в точках 0 и 1, чтобы наклон в точках 0 и 2 был равен 0, и первая производная была непрерывной. В совокупности это дает семь ограничений, в то время как неизвестных параметров восемь и, следовательно, нужно еще одно условие для однозначного определения интерполяционного ядра. В частности, если интерполяционную функцию привести в соответствие с первыми тремя членами ее разложения в ряд Тейлора, то неизвестный параметр  $a$  должен быть равен  $\left(-\frac{1}{2}\right)$ . Для практических

задач удобнее все семь коэффициентов определить через неизвестный параметр  $a$  и интерполяционное ядро представлять в виде [5.10]

$$g(s) = \begin{cases} (a+2)|s|^3 - (a+3)|s|^2 + 1, & |s| \leq 1; \\ a|s|^3 - 5a|s|^2 + 8a|s| - 4a, & 1 < |s| \leq 2; \\ 0, & |s| > 2. \end{cases} \quad (5.24)$$

Когда константа  $a$  отрицательная, ядро (5.24) положительное в интервале от 0 до 1 и отрицательное в интервале от 1 до 2. Когда  $a$  возрастает, глубина боковых лепестков в интервале от 1 до 2 также увеличивается. Таким образом, с отрицательным значением свободной константы  $a$  интерполяционное ядро имеет вид усеченной sinc-функции. Выяснилось [5.11], что эта функция имеет более предпочтительные высокочастотные свойства, нежели кубический B-сплайн, и было предложено называть данную функцию *высокоразрешающим интерполяционным кубическим сплайном*. Варьируя значение параметра  $a$  в пределах от  $-1$  (рис. 5.11, а) до  $-1/2$  (рис. 5.11, б), в каждом конкретном случае можно добиться приемлемой точности при передискретизации.

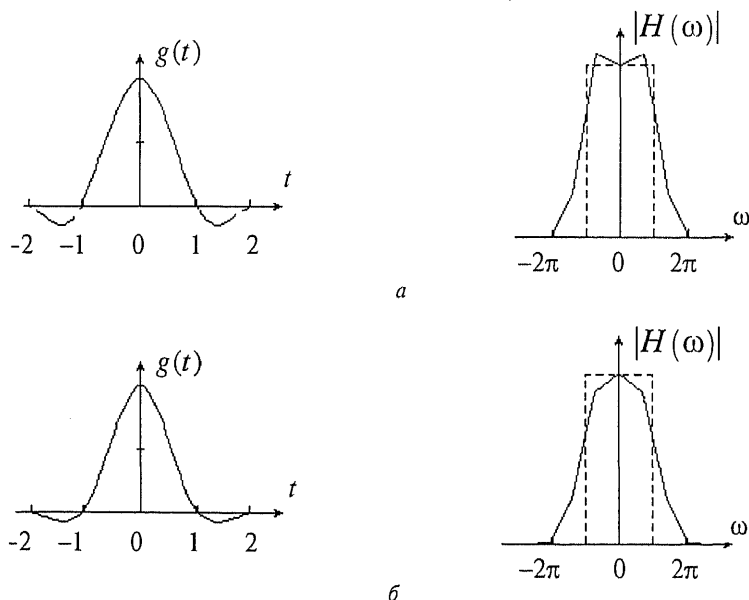


Рис. 5.11. Высокоразрешающий интерполяционный кубический сплайн и его частотная характеристика:  $a = -1$  (а);  $a = -0.5$  (б)

### 5.3. ПРИВЯЗКА ИЗОБРАЖЕНИЙ

В практике обработки изображений задача поиска соответствия получила большое распространение и известна как проблема «поиска по образцу». Формально ее можно рассматривать как процесс отождествления эталонного изображения (образа фрагмента) на первом снимке с одним из множества образов фрагментов, лежащих в некоторой (задаваемой) области (зоне поиска) второго снимка. Алгоритмы установления сходства в своих основополагающих вариантах в той или иной степени связаны с получением характеристик стохастической взаимосвязи сравниваемых фрагментов изображений [5.2, гл.19]. Все они основываются на идеях корреляционной и спектральной теории сигналов, и для соответствующих критериев получены экспериментальные характеристики основных процедур поиска по образцу.

#### 5.3.1. КОРРЕЛЯЦИОННЫЙ КРИТЕРИЙ СХОДСТВА

Будем считать, что изображение эталонного фрагмента (выбранного на снимке  $A$  и представляемого матрицей  $U_0$  размером  $n \times n$ ), сравнивается с изображениями фрагментов снимка  $B$  в «зоне поиска»  $\Omega$  размером  $L \times L$ ,  $L = n + m$ . Перекрытие между фрагментами определяется шагом  $h$  дискретной решетки  $hZ^2$  (в плоскости  $P^2$ ), на которой заданы наблюдаемые переменные  $\{u_0(x), x = (x, y)\}$  на  $A$  или  $\{u(x)\}$  на  $B$ . В процессе скользящего поиска (когда каждый очередной фрагмент получается из предыдущего простым сдвигом на один шаг) вычисляется «функция сходства» между изображением эталонного фрагмента  $\{u_0(x), x \in \Gamma_A\}$  и изображениями текущих (контролируемых) фрагментов  $\{u(x), x \in \Gamma_B\}$ . Здесь требуется найти функцию сходства, которая бы с максимально возможной точностью и достоверностью позволяла локализовать фрагмент, соответствующий изображению эталонного фрагмента, фиксируя таким образом сопряженные точки на снимках.

Взаимно соответствующие элементы изображений одного объекта на снимках должны, очевидно, удовлетворять соотношению

$$u_0(x, y) = (au(x + k, y + l) + b) \operatorname{rect}\left(\frac{x}{n}, \frac{y}{n}\right) + \varepsilon(x, y), \quad (5.25)$$

где  $a$  и  $b$  — параметры контраста и средней освещенности;  $k, l$  — параметры относительного сдвига образца и его аналога на контролируемом снимке;  $\varepsilon(x, y)$  — шум;

$$\text{rect}\left(\frac{x}{n}, \frac{y}{n}\right) = \begin{cases} 1, & x \leq n, y \leq n; \\ 0, & \text{иначе.} \end{cases}$$

В такой формулировке процедура селекции образца должна найти параметры  $k$  и  $l$ , характеризующие сдвиг реперных фрагментов.

Ради простоты будем считать, что параметр  $b$  не меняется по полю снимков, что позволяет перейти к центрированным переменным

$$\bar{u}(x, y) = u(x, y) - \bar{u}, \bar{u} = \frac{1}{L^2} \sum_{(x, y) \in \Omega} u(x, y);$$

$$\bar{u}_0(x, y) = u_0(x, y) - \bar{u}_0, \bar{u}_0 = \frac{1}{n^2} \sum_{x, y=1}^n u_0(x, y).$$

В качестве меры различия в точке  $(k, l)$  будем брать средне-квадратическую ошибку

$$\epsilon_a^2(k, l) = \sum_x \sum_y [\bar{u}_0(x, y) - a\bar{u}(x + k, y + l)]^2, \quad (5.26)$$

которая минимизируется перебором всех допускаемых сдвигов эталона по заданной области контролируемого снимка. Считается, что в точке экстремума реализуется сходство, если  $\epsilon_a^2(k, l) \leq \lambda$ , где  $\lambda$  — некоторый установленный порог. Из требования минимума ошибки  $(\epsilon_a^2(k, l))'_a = 0$  находим оценку  $a$ , подставляем ее в формулу (5.26) и приходим к выражению

$$\epsilon_a^2(k, l) = \sum_x \sum_y [\bar{u}_0(x, y)]^2 - \frac{\left[ \sum_x \sum_y \bar{u}_0(x, y) \bar{u}(x, y) \right]^2}{\sum_x \sum_y [\bar{u}(x, y)]^2}. \quad (5.27)$$

Первый член выражения (5.27) — «энергия» эталонного сигнала — является величиной постоянной, не зависящей от параметров сдвига  $(k, l)$ . Поэтому точка экстремума не изменится, если мы будем нормировать среднеквадратическую ошибку к энергии эталона

$$\bar{\epsilon}_a^2(k, l) = 1 - \frac{\left[ \sum_x \sum_y \bar{u}_0(x, y) \bar{u}(x, y) \right]^2}{\sum_x \sum_y [\bar{u}(x, y)]^2 \sum_x \sum_y [\bar{u}_0(x, y)]^2},$$

и вместо минимума нормированной среднеквадратической ошибки будем искать максимум *коэффициента корреляции* текущего фрагмента с эталоном

$$\hat{r}(k, l) = \frac{\left[ \sum_x \sum_y \bar{u}_0(x, y) \bar{u}(x, y) \right]}{\left\{ \sum_x \sum_y [\bar{u}(x, y)]^2 \sum_x \sum_y [\bar{u}_0(x, y)]^2 \right\}^{\frac{1}{2}}}. \quad (5.28)$$

Соблюдение условий достоверности обнаружения также приводит к необходимости устанавливать порог для величины взаимной корреляции  $\max \hat{r}(k, l)$ : если  $\max \hat{r}(k, l) \geq r_{\text{пор}}$ , то с заданной вероятностью гарантируется действительное сходство найденной пары фрагментов. Величина порога определяется функцией распределения коэффициента корреляции (при случайных выборках) и задаваемой доверительной вероятностью принятия решения о действительном сходстве фрагментов.

Функционирование данного (по существу классического) алгоритма при наличии искажений в изображениях рассмотрено в работе [5.9]. Различия между эталонным и текущим ( $u_0(x, y)$  и  $u(x, y)$ ) изображениями были обусловлены аддитивным шумом и геометрическими искажениями, которые моделировались аффинными преобразованиями координат изображений:

$$u_0(\mathbf{x}) = u(\mathbf{A}\mathbf{x} + \mathbf{t}),$$

где  $\mathbf{x} = (x, y)$ ;  $\mathbf{A} = \alpha \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$  — матрица относительного поворота изображений на угол  $\theta$ ;  $\alpha$  — коэффициент изменения масштаба. В работе показано, что среднее значение основного пика корреляционной функции геометрически искаженных изображений, нормированное к средней величине пика при отсутствии искажений, зависит от интенсивности искажений и при малых  $\theta$  и  $|1 - \alpha|$  имеет вид

$$d \approx \sqrt{(1 - \alpha)^2 + \theta^2}.$$

Было практически продемонстрировано, что серьезным недостатком корреляционной меры сходства является ее чувстви-

тельность к геометрическим искажениям видимых размеров сопряженных фрагментов при изменении ракурса съемки.

Обычно за критерии эффективности процедур идентификации сходства принимают точность совмещения фрагментов и вероятность ложной привязки, когда экстремум функционала сходства значимо смещен относительно истинного положения. Анализ результатов имитационных экспериментов позволил сделать следующие выводы.

При наличии геометрических искажений существует оптимальный размер фрагмента эталонного изображения, позволяющий минимизировать вероятность ложной привязки. Оптимальный размер фрагмента пропорционален эффективному радиусу корреляции (полуширине графика автокорреляционной функции) и уменьшается с увеличением геометрических искажений.

При заданном уровне искажений размер эталонного изображения, при котором погрешность совмещения минимальна, меньше, чем размер изображения, необходимый для минимизации вероятности ложной привязки.

Здесь можно порекомендовать использовать полезную модификацию метода идентификации сходства, заключающуюся в том, что искажения геометрии на втором снимке (относительно первого) предварительно компенсируются аффинной (или полиномиальной) «подгонкой». Например, параметры аффинного преобразования

$$A = \begin{bmatrix} a & b \\ d & e \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} g \\ h \end{bmatrix}$$

можно оценивать адаптивно (в несколько «проходов»), когда на первом этапе задается достаточно большая зона поиска по образцу, что позволяет на искаженном (по отношению к исходному) снимке находить сопряженные точки. Даже трех пар опорных точек достаточно, чтобы оценить (в первом приближении) параметры аффинного преобразования и осуществить аффинную подгонку геометрии изображения  $u(\mathbf{x})$  к геометрии изображения  $u_0(\mathbf{x})$ :  $u_0(\mathbf{x}) \approx u(A\mathbf{x} + \mathbf{a})$ . Это дает возможность повторным просмотром найти уже существенно большее число пар сопряженных точек на исходном  $u_0(\mathbf{x})$  и аффинно-преобразованном  $u(A^{(2)}\mathbf{x} + \mathbf{a}^{(2)})$  снимках и уточнить по ним параметры аффинной аппроксимации. Дальнейшее повторение этой процедуры позволяет, в принципе, идентифицировать любое (допустимое данной

аппроксимацией) число пар сопряженных точек и, следовательно, добиться заданной точности в оценивании параметров геометрического преобразования.

Поиск по образцу в данном методе сводится к вычислению нормированной взаимной корреляции распределения яркости (двумерного сигнала) на текущем фрагменте первого снимка с распределениями яркостей фрагментов, лежащих в некоторой предполагаемой окрестности образа этого фрагмента, на аффинно-преобразованном втором снимке и определению целочисленных параметров взаимного смещения исходного фрагмента и его образа, устанавливаемого по экстремуму корреляционного функционала.

### 5.3.2. ЛОКАЛЬНОЕ УТОЧНЕНИЕ СДВИГА

После определения целочисленных параметров смещения чаще всего требуется локальное уточнение сдвига фрагментов в пределах дискрета (шага целочисленной решетки) [5.8, гл.15]. Чтобы найти соответствующий вектор сдвига  $(\delta x, \delta y)$ , будем считать, что кросс-корреляционная функция  $\hat{r}_{0p}(x, y)$  в окрестности точки экстремума  $(\hat{x}, \hat{y})$  разлагается в ряд Тейлора:

$$\begin{aligned} \hat{r}_{0p}(x, y) \approx \hat{r}_{0p}(\hat{x}, \hat{y}) + \frac{1}{2} \frac{\partial^2 \hat{r}_{0p}}{\partial x^2} \Big|_{\hat{x}, \hat{y}} (x - \hat{x})^2 + \\ + \frac{1}{2} \frac{\partial^2 \hat{r}_{0p}}{\partial y^2} \Big|_{\hat{x}, \hat{y}} (y - \hat{y})^2 + \frac{\partial^2 \hat{r}_{0p}}{\partial x \partial y} (x - \hat{x})(y - \hat{y}) \end{aligned} \quad (5.29)$$

(первые производные в точке экстремума равны нулю). Дифференцируя данное выражение, непосредственно получаем

$$\begin{aligned} \frac{\partial \hat{r}_{0p}}{\partial x} \Big|_{\hat{x}, \hat{y}} = -\frac{\partial^2 \hat{r}_{0p}}{\partial x^2} \Big|_{\hat{x}, \hat{y}} \delta x - \frac{\partial^2 \hat{r}_{0p}}{\partial x \partial y} \Big|_{\hat{x}, \hat{y}} \delta y, \\ \frac{\partial \hat{r}_{0p}}{\partial y} \Big|_{\hat{x}, \hat{y}} = -\frac{\partial^2 \hat{r}_{0p}}{\partial y^2} \Big|_{\hat{x}, \hat{y}} \delta y - \frac{\partial^2 \hat{r}_{0p}}{\partial x \partial y} \Big|_{\hat{x}, \hat{y}} \delta x. \end{aligned}$$

В результате приходим к матричному уравнению для параметров  $(\delta x, \delta y)$ :

$$\begin{bmatrix} (\hat{r}_{0p})''_{xx} & (\hat{r}_{0p})''_{xy} \\ (\hat{r}_{0p})''_{yx} & (\hat{r}_{0p})''_{yy} \end{bmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = - \begin{pmatrix} (\hat{r}_{0p})'_x \\ (\hat{r}_{0p})'_y \end{pmatrix}. \quad (5.30)$$

Таким образом, в результате всех вычислений, проведенных над элементами изображений пары снимков, координаты пар сопряженных точек  $x_1$ ,  $x_2$  в целом связываются соотношением

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \hat{\mathbf{A}} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} + \begin{pmatrix} \delta \hat{x} \\ \delta \hat{y} \end{pmatrix} + \hat{\mathbf{a}}. \quad (5.31)$$

### 5.3.3. КРОСС-СПЕКТРАЛЬНАЯ МЕРА СХОДСТВА

Здесь мы разовьем далее охарактеризованный метод решения задачи привязки. Для этого будем исходить из исследования обобщенной кросс-спектральной меры сходства. Для единообразия запишем матрицу эталонного фрагмента в левом верхнем углу нулевой матрицы порядка  $L \times L$ . Это позволяет перейти от централизованных переменных  $\tilde{u}_0(x, y), \tilde{u}(x, y)$  к спектральным компонентам (Фурье-образам) изображений  $\tilde{\mathbf{U}}, \tilde{\mathbf{U}}_0$ :

$$s_{pq}(\tilde{\mathbf{U}}) = \{ F(\tilde{\mathbf{U}}) \}_{pq} = |s_{pq}(\tilde{\mathbf{U}})| \exp(i\varphi_{pq}(\tilde{\mathbf{U}})), \quad p, q = 0, 1, \dots, L-1;$$

$$|s_{pq}(\tilde{\mathbf{U}})| = \sqrt{\text{Re}^2 s_{pq}(\tilde{\mathbf{U}}) + \text{Im}^2 s_{pq}(\tilde{\mathbf{U}})}, \quad \varphi_{pq}(\tilde{\mathbf{U}}) = \arctg \left( \frac{\text{Im} s_{pq}(\tilde{\mathbf{U}})}{\text{Re} s_{pq}(\tilde{\mathbf{U}})} \right).$$

Поскольку при смещении «образца»  $\{\tilde{u}_0(x, y), x \leq n, y \leq n\}$  в области поиска  $\Omega$  меняется только его фазовый спектр

$$\{ F(\tilde{\mathbf{U}}_0(x+k, y+l)) \}_{pq} = s_{pq}(\tilde{\mathbf{U}}_0) \exp \left( i \frac{2\pi}{L} (pk + ql) \right),$$

будем минимизировать функционал

$$\begin{aligned} F_{kl} = & \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} h_{pq} \left( s_{pq}(\tilde{\mathbf{U}}) - e^{i \frac{2\pi}{L} (pk+ql)} s_{pq}(\tilde{\mathbf{U}}_0) \right)^* \times \\ & \times \left( s_{pq}(\tilde{\mathbf{U}}) - e^{i \frac{2\pi}{L} (pk+ql)} s_{pq}(\tilde{\mathbf{U}}_0) \right) \end{aligned} \quad (5.32)$$



по параметрам сдвига  $k, l$ , где звездочкой обозначена операция комплексного сопряжения, а функция  $h_{pq}$  осуществляет «взвешивание» разностной меры спектральных компонентов. Так как «энергия» изображений фрагментов  $|s_{pq}(\bar{U})|^2$ ,  $|s_{pq}(\bar{U}_0)|^2$  не зависит от параметров  $k, l$ , то положение минимума функционала  $F_{kl}$  соответствует точке экстремума перекрестного члена при раскрытии скобок в (5.32):

$$C_{kl}(\bar{U}, \bar{U}_0) = \sum_{p=0}^{L-1} \sum_{q=0}^{L-1} h_{pq} s_{pq}^*(\bar{U}_0) s_{pq}(\bar{U}) \exp\left(i \frac{2\pi}{L} (pk + ql)\right). \quad (5.33)$$

Точность локализации определяется крутизной данной кросс-спектральной функции вблизи экстремума, характеризующего положение образа эталонного фрагмента. С этой точки зрения наилучшей моделью для  $C_{kl}(\bar{U}, \bar{U}_0)$  служит дискретный вариант дельта функции  $\delta(k - \hat{k}, l - \hat{l})$ . Нетрудно убедиться, что последнему условию удовлетворяет весовая функция  $h_{pq}$  вида

$$\hat{h}_{pq} = |s_{pq}^*(\bar{U}_0) s_{pq}(\bar{U})|^{-1}, \quad (5.34)$$

так как в этом случае обратное дискретное фурье-преобразование

$$\begin{aligned} F^{-1} \left\{ \exp \left[ i(\varphi_{pq}(\bar{U}(x+k, y+l)) - \varphi_{pq}(\bar{U}_0(x, y))) \right] \right\} = \\ = \delta(k - \hat{k}, l - \hat{l}) = \begin{cases} 1, k = \hat{k}, l = \hat{l}; \\ 0, (k, l) \neq (\hat{k}, \hat{l}). \end{cases} \end{aligned}$$

Полученный результат можно интерпретировать следующим образом. Обнаружитель с передаточной функцией (5.34) представляет собой «фазовый» фильтр, т.е. фильтр, который фазовую часть комплексного спектра изображений оставляет без изменений, а амплитудный спектр нормализует. Нетрудно усмотреть, что в данном варианте кросс-спектральная мера  $C_{kl}(\bar{U}, \bar{U}_0)$  есть Фурье-образ функции когерентности

$$\gamma_{pq} = \frac{s_{pq}^*(\bar{U}_0) s_{pq}(\bar{U})}{|s_{pq}^*(\bar{U}_0) s_{pq}(\bar{U})|}.$$

Тем самым можно отнести все признаки оптимальности когерентного функционала [5.6] и к кросс-спектральной мере сходства, допускающей к тому же эффективную реализацию на основе быстрого преобразования Фурье.

Следует отметить, что выражение (5.33) при  $h_{pq} \equiv 1$  соответствует классическому корреляционному алгоритму, реализуемому в спектральной области. Функционирование данного алгоритма при наличии геометрических искажений и выводы, сделанные на основе моделирования, выше были нами приведены.

Анализ результатов имитационных экспериментов показал [5.7], что аналогичные выводы справедливы и для когерентного (фазового) алгоритма. Однако рабочая зона идентификации фазового фильтра (по величине допустимых геометрических искажений) меньше аналогичной зоны для корреляционного алгоритма. Это и понятно, так как чем меньше эффективный радиус кросс-корреляционной функции сходства  $C_{kl}(\vec{U}, \vec{U}_0)$ , тем более чувствителен алгоритм к изменению геометрии идентифицируемых фрагментов. В частности, алгоритм нормированной корреляции ведет удовлетворительный поиск по образцу для снимков, развернутых на угол до  $10^\circ$  ( $n = 18$ ,  $L = 25$ ). Алгоритм фазовой корреляции здесь нормально функционирует, если угол поворота не превышает  $5^\circ$ . На рис. 5.12, *а* представлено типичное поведение кросс-спектральной меры сходства в рассматриваемых вариантах, а на рис. 5.12, *б* даны графики оцененных вероятностей

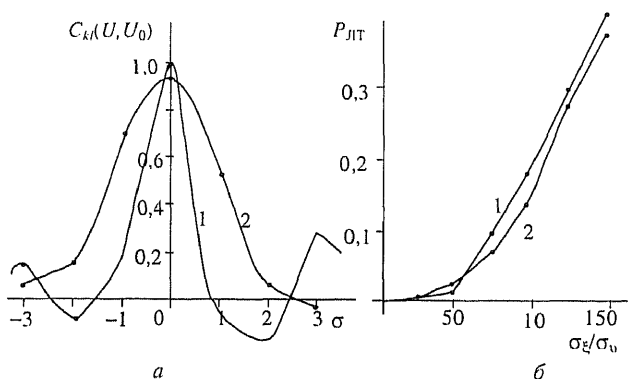


Рис. 5.12. Меры эффективности алгоритмов идентификации сходства: *а* — поведение кросс-спектральной меры сходства вблизи ее экстремума; 1 — отклик фазового фильтра, 2 — отклик корреляционного обнаружителя; *б* — оценки вероятностей ложной привязки (ВЛП) в зависимости от отношения амплитуды шума  $\sigma_z$  к амплитуде фонового сигнала (в %): 1 — ВЛП в фазовом алгоритме, 2 — в корреляционном алгоритме

ложной привязки в зависимости от уровня случайной составляющей сигнала на контролируемом снимке. Видно, что фазовый фильтр в достаточно широком диапазоне мощностей помехи (до 55 %) имеет меньшую вероятность ложной привязки фрагментов, нежели классический корреляционный алгоритм. Для иллюстрации изложенных методов на рис. 5.14 показан результат привязки радиолокационных снимков (рис. 5.13).

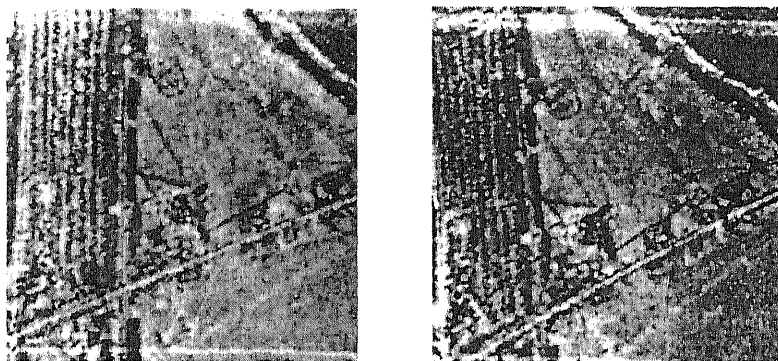


Рис. 5.13. Радиолокационные изображения местности, снятые при двух пролетах самолета (разные ракурсы съемки)

#### 5.3.4. ПРИВЯЗКА ПО ЛОКАЛЬНЫМ НЕОДНОРОДНОСТЯМ

Если в некоторой части изображения уровень яркости более или менее постоянен или на одном из снимков искажен ино-родными включениями, то сопряженные точки искать достаточно трудно. Кросскорреляционные методы, использующие фрагменты меньшие, чем данная область однородной яркости, не дадут ярко выраженного максимума либо максимум не превысит заданный пороговый уровень.

На левом снимке рис. 5.14 большие светлые рамки соответствуют зонам поиска для правого снимка рис. 5.13, малые рамки характеризуют размер эталонных фрагментов, выбираемых на левом снимке рис. 5.13 (по изображениям соответствующих фрагментов вычислялась кросс-спектральная мера сходства). Смещение малых рамок от центра больших рамок указывает на величину найденных локальных геометрических деформаций между снимками. Правый снимок соответствует правому снимку на рис. 5.13, геометрически преобразованному в формат левого снимка (полиномиальное преобразование третьей степени с па-

раметрами, оценки которых мало отличаются от аффинного преобразования; передискретизация осуществлялась на основе высокоразрешающей бикубической сплайн-интерполяции). Это достаточно хорошо видно на рис. 5.14, где рамками обозначены фрагменты, для которых взаимная мера сходства между изображениями находится в доверительном интервале.

В то же время в большей части областей снимков не регистрировалось действительного сходства в выбранных фрагментах зоны поиска. Поэтому может оказаться более разумным осуществлять поиск реперных фрагментов только в информационно насыщенных областях, где яркость быстро меняется, например, на краях между более или менее однородными областями.

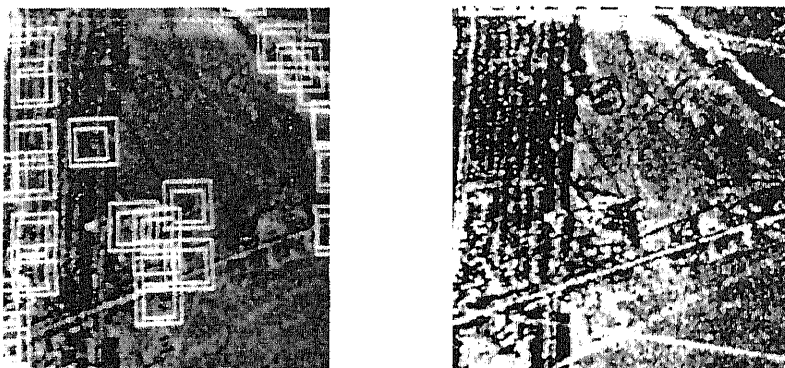


Рис. 5.14. Результаты привязки изображений

Выделение краев [5.2, гл.17] можно рассматривать как средство быстрого предварительного просмотра эталонного изображения и отбора информативных областей. В полученных областях далее выбирают эталонные реперные фрагменты и осуществляется поиск корреляционными методами соответствующих им фрагментов на контролируемых изображениях. В этом случае привязка будет осуществляться по заведомо информативным областям, что повышает точность локализации сопряженных точек и уменьшает вероятность ложного отождествления.

## ВОПРОСЫ К ГЛАВЕ 5

5.1. Охарактеризуйте круг проблем, решение которых приводит к привязке последовательности изображений и их взаимной геометрической коррекции.

5.2. Выстройте иерархию геометрических преобразований. Какие из них сохраняют параллельность прямых?

5.3. Постройте матрицу поворота (в однородных координатах) вокруг точки плоскости  $(x_0, y_0)$  на угол  $\varphi$  (указание: совместить центр поворота с началом координат, повернуть, вернуть центр поворота в прежнее положение и перемножить полученные матрицы элементарных преобразований).

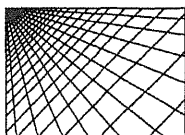
5.4. Чем отличается аффинная плоскость от евклидовой плоскости и от проективной плоскости?

5.5. В каких ситуациях рекомендуется применять полиномиальную аппроксимацию (в том числе и полиномы Чебышева) для описания геометрических деформаций?

5.6. Опишите процедуру восстановления изображения в преобразованных координатах. Какие используются методы интерполяции и в чем их различие?

5.7. Назовите меры сходства изображений и охарактеризуйте их относительные свойства (смысл их оптимальности, точность локализации, достоверность, устойчивость к геометрическим деформациям).

5.8. Опишите процедуру привязки изображений и методы ее ускорения.



## ГЛАВА 6

### ФОТОГРАММЕТРИЯ И СТЕРЕОВИДЕНИЕ

В обычном изображении трехмерного пространства информация о расстоянии до различных элементов сцены проявляется только в виде косвенных признаков: через относительные размеры объектов, затенение одних объектов другими, различную освещенность и т.д. Один из способов получения информации о глубине состоит в регистрации нескольких изображений сцены под различными ракурсами. В этом случае точки сцены дают изображения, относительное положение которых зависит от расстояния до точки наблюдения. Оказывается, сопоставляя эти изображения, в ряде случаев можно реконструировать трехмерную структуру сцены. Основы теории определения положения объектов в пространстве по их перспективным изображениям были разработаны еще в средние века, а в XVIII веке перспективные рисунки стали использовать в топографических целях. Появление фотографии положило начало фотограмметрии — науке об определении формы, размеров и пространственного положения различных объектов посредством измерения их фотографических изображений. В начале XX века был изобретен стереокомпаратор — прибор для измерения пространственного положения объектов по паре перекрывающихся фотографических изображений. Интерес к методам восстановления трехмерной структуры сцен по их плоским изображениям возобновился в середине XX столетия в связи с исследованиями в области искусственного интеллекта, а практическая потребность в робототехнических устройствах, способных ориентироваться в трехмерном пространстве, постоянно поддерживает этот интерес в последние десятилетия.

В этой главе мы рассмотрим процесс формирования изображений, соотношения между координатами точек сцены и их изображениями и методы оценивания параметров системы регистрации и трехмерной структуры сцены.

### 6.1. МОДЕЛЬ РЕГИСТРИРУЮЩЕЙ КАМЕРЫ

Строго говоря, различные точки пространства предметов отображаются оптической системой камеры в пространстве изображений на различных расстояниях от фокальной плоскости. Однако, если расстояние между камерой и наблюдаемой сценой значительно превышает фокусное расстояние оптической системы, можно считать, что изображение строится в ее фокальной плоскости. В этом случае можно использовать проективную модель камеры, в которой изображение трехмерного объекта получается проектированием его в фокальную плоскость (плоскость изображения) через единственную точку, называемую оптическим центром. Прямая линия, перпендикулярная плоскости изображения и проходящая через эту точку, называется оптической осью камеры, а точка пересечения оптической оси с плоскостью изображения — главной точкой.

Определим в трехмерном пространстве ортогональную правую систему координат  $OXYZ$ , начало которой совпадает с оптическим центром, ось  $OZ$  — с оптической осью камеры. Такая система называется *стандартной системой координат* камеры. Пусть плоскость изображения находится на расстоянии  $f$  от оптического центра. В этой плоскости зададим систему координат  $oxy$  с началом в главной точке и осями  $ox$  и  $oy$ , параллельными осям  $OX$  и  $OY$  соответственно (рис. 6.1). Легко убедиться, что в стандартной системе координат проекцией точки трехмерного пространства  $M$  с координатами  $(X, Y, Z)$  является точка  $m$  в плоскости изображения с координатами  $(x, y)$ , причем

$$x = fX / Z, \quad y = fY / Z.$$

Для полного описания камеры следует учесть, что для регистрации изображения в плоскости изображения камеры имеется какой-либо фотоприемник. В фотокамере это фотопластинка или фотопленка, в видеокамере это видикон или фотоприемная полупроводниковая матрица, в некоторых специальных камерах — электромеханический сканер. В общем случае координаты в фотоприемнике измеряют в единицах, отличных от единиц, задающих координаты в стандартной системе. Например, если ис-

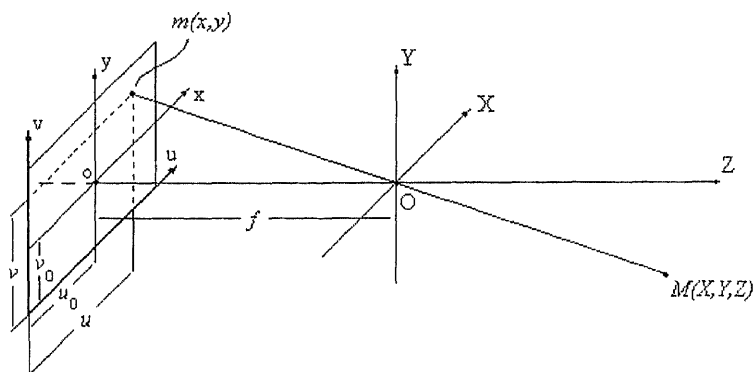


Рис. 6.1. Система координат проективной камеры

пользуется матричный фотоприемник, его естественными координатами являются номер строки и номер столбца фотоприемной ячейки. Поэтому для полного описания камеры необходимо выразить координаты точки  $m$  в естественных единицах фотоприемника. В достаточно общем для любых фотоприемников виде (рис. 6.1) это может выглядеть как

$$u = x / w + u_0, \quad v = y / h + v_0,$$

где  $(u_0, v_0)$  — координаты главной точки относительно начала координат фотоприемника (в естественных координатах фотоприемника);  $w$  и  $h$  — масштабы вдоль осей  $ox$  и  $oy$  (например, расстояния между ячейками матричного фотоприемника вдоль строк и столбцов).

В новой системе координаты проекции точки  $m$  примут вид

$$u = \frac{fX}{wZ} + u_0, \quad v = \frac{fY}{hZ} + v_0. \quad (6.1)$$

Для последующего изложения введем трехмерный вектор  $\mathbf{M} = (X, Y, Z)^T$ , соответствующий точке  $M$ , и двумерный вектор  $\mathbf{m} = (x, y)^T$ , соответствующий точке  $m$ . Определим также вектор однородных (см. гл. 5) внутренних координат камеры  $\mathbf{v} = (u, v, 1)^T$ . Используя эти обозначения, соотношения (6.1) можно представить в компактной векторно-матричной записи:

$$\mathbf{Zv} = \mathbf{AM}, \quad (6.2)$$



где  $A = \begin{bmatrix} f/w & 0 & u_0 \\ 0 & f/h & v_0 \\ 0 & 0 & 1 \end{bmatrix}$  — матрица, известная под названием

матрицы *внутренних параметров* камеры, поскольку она содержит только параметры оптической системы и фотоприемника камеры.

## 6.2. СВЯЗЬ МЕЖДУ РАЗЛИЧНЫМИ СИСТЕМАМИ КООРДИНАТ

В общем случае трехмерные координаты точки могут быть заданы в системе, не совпадающей со стандартной системой координат камеры (назовем ее *глобальной*). Пусть  $OXYZ$  — глобальная система координат, а  $O'X'Y'Z'$  — стандартная система координат камеры. Переход от системы  $OXYZ$  к системе  $O'X'Y'Z'$  можно осуществить поворотом координатных осей к системе  $OX''Y''Z''$  и последующим смещением начала координат. Тогда связь между координатами точки  $M$  в глобальной и стандартной системах может быть представлена как

$$M' = RM + t, \quad (6.3)$$

где  $M$  и  $M'$  — векторы пространственных координат точки  $M$  в глобальной и стандартной системах соответственно;  $R$  — матрица размером  $3 \times 3$ , описывающая поворот стандартной системы координат относительно глобальной; компонентами матрицы являются направляющие косинусы осей глобальной системы в стандартной системе координат;  $t$  — трехмерный вектор смещения начала координат глобальной системы относительно начала координат стандартной.

На рис. 6.2 схематически показано преобразование координат. Здесь  $\alpha$ ,  $\beta$  и  $\gamma$  — углы, образованные осью  $OX''$  с осями  $OX$ ,  $OY$  и  $OZ$  соответственно. Элементы первой строки матрицы  $R$  [6.1, п. 14.10] содержат косинусы этих углов:  $r_{11} = \cos \alpha$ ,  $r_{12} = \cos \beta$ ,

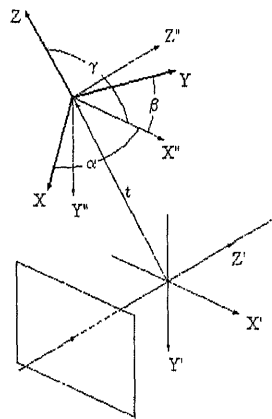


Рис. 6.2. Переход от глобальной системы координат к стандартной системе координат камеры

$r_{13} = \cos \gamma$ . Аналогично, вторая и третья строки матрицы содержат косинусы углов, образованных соответственно осями  $O'Y'$  и  $O'Z'$  с осями глобальной системы координат.

Особенность матрицы  $\mathbf{R}$  состоит в том, что она зависит только от трех параметров, поскольку все девять ее элементов связаны шестью уравнениями связи и, следовательно, не являются независимыми. Обозначив строки матрицы в виде векторов  $\mathbf{r}_1^T$ ,  $\mathbf{r}_2^T$  и  $\mathbf{r}_3^T$ , эти уравнения можно представить так:

$$\mathbf{r}_1^T \mathbf{r}_1 = 1, \mathbf{r}_2^T \mathbf{r}_2 = 1, \mathbf{r}_3^T \mathbf{r}_3 = 1, \mathbf{r}_1^T \mathbf{r}_2 = 0, \mathbf{r}_1^T \mathbf{r}_3 = 0, \mathbf{r}_2^T \mathbf{r}_3 = 0. \quad (6.4)$$

Уравнения (6.4) являются условиями взаимной ортогональности векторов  $\mathbf{r}_i$ . Матрица, построенная из таких векторов, называется *ортогональной*. Для ортогональной матрицы справедливо соотношение  $\mathbf{R}^{-1} = \mathbf{R}^T$ . Условие взаимной ортогональности векторов  $\mathbf{r}_i$  в трехмерном пространстве можно выразить в другой удобной форме, которая понадобится нам позже:

$$\mathbf{Lc} = 0, \mathbf{r}_3 = \pm(\mathbf{r}_1 \times \mathbf{r}_2), \mathbf{r}_2 = \pm(\mathbf{r}_3 \times \mathbf{r}_1), \mathbf{r}_1 = \pm(\mathbf{r}_2 \times \mathbf{r}_3). \quad (6.5)$$

Верхний знак соответствует случаю, когда матрица  $\mathbf{R}$  представляет преобразование, не изменяющее взаимной ориентации осей системы, а нижний — преобразование, изменяющее правую систему координат на левую и наоборот. Смысл вектора  $\mathbf{t} = (t_x, t_y, t_z)^T$  ясен непосредственно из рисунка.

### 6.3. СТЕРЕОСКОПИЧЕСКАЯ СИСТЕМА

Рассмотрим ситуацию, когда две камеры, находящиеся в разных точках, регистрируют одну и ту же сцену. Пара изображений, получаемых при этом, называется *стереопарой*. Обратимся сначала к простейшему случаю. Пусть одинаковые камеры расположены так, что их оптические оси параллельны, а прямая, проходящая через оптические центры, перпендикулярна оптическим осям (эта прямая называется *базовой линией*, а ее отрезок, заключенный между оптическими центрами, — *базой*). Положим длину базы равной  $b$ . Выберем такую глобальную систему координат, начало которой  $O$  расположено на базовой линии посередине между оптическими центрами камер, ось  $OZ$  параллельна оптическим осям, а ось  $OX$  направлена вдоль базовой линии (рис. 6.3). Пусть начала координат в плоскостях изображений камер совпадают с главными точками ( $u_0 = v_0 = 0$ ), а единицы

измерения координат в глобальной системе и в плоскостях изображения камер одинаковы ( $w = h = 1$ ).

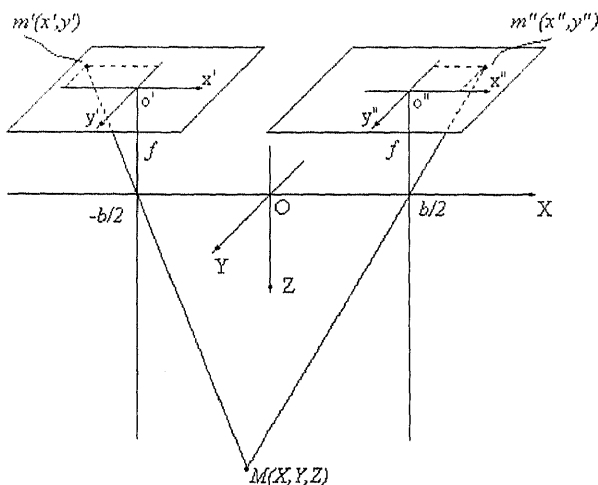


Рис. 6.5. Простейшая стереоскопическая система

Выберем точку  $M$  с глобальными координатами  $(X, Y, Z)$ . Координаты ее проекции в плоскости изображения первой (левой) камеры обозначим через  $(x', y')$ , а в плоскости изображения второй (правой) камеры — через  $(x'', y'')$ . (Проекция одной и той же точки  $M$  в плоскостях изображений разных камер называются *сопряженными точками*.) Нетрудно проверить, что

$$x' = f(X + b/2) / Z, \quad x'' = f(X - b/2) / Z, \quad y' = y'' = fY / Z.$$

Заметим, что в направлении, перпендикулярном направлению базовой линии, координаты сопряженных точек ( $y$ -координаты) совпадают. Это обстоятельство имеет большое значение при автоматизированном поиске сопряженных точек на стереопаре, позволяя существенно сократить размеры зоны поиска. Из первых двух соотношений следует, что

$$Z = fb / (x' - x''). \quad (6.6)$$

Это означает, что, зная геометрию съемки и выполнив измерения координат проекций одной и той же точки в плоскостях изображения камер, можно вычислить глубину (координату  $Z$ )

этой точки. Более того, полученные соотношения позволяют вычислить полностью трехмерные координаты точки:

$$X = b \frac{(x' + x'')}{2(x' - x'')}, Y = b \frac{(y' + y'')}{2(x' - x'')}. \quad (6.7)$$

Разность  $d = x' - x''$  называется *диспарантностью*. Из (6.6) и (6.7) видно, что ошибки в координатах проекций сильнее сказываются при малой диспарантности и, следовательно, расстояния до далеких объектов измеряются менее точно, чем до близких. С другой стороны, при фиксированной дальности диспарантность пропорциональна размеру базы, поэтому точность измерений повышается с увеличением базы. Далее мы, однако, увидим, что увеличение базы может привести к ошибкам, которые не компенсируются повышением точности измерений.

Теперь рассмотрим общий случай, когда оптические оси камер не параллельны и направление смещения оптического центра одной камеры относительно оптического центра другой произвольно (рис. 6.4). Введем для каждой камеры свою стандартную систему координат так, как это было сделано в разделе 6.1. Пусть первой камере соответствует система координат  $O'X'Y'Z'$ , а второй —  $O''X''Y''Z''$  (рис. 6.4). Пусть вектор  $M' = (X', Y', Z')^T$  характеризует координаты некоторой точки  $M$  трехмерного пространства в системе первой камеры, а вектор  $M'' = (X'', Y'', Z'')^T$  — в системе второй. Переход от глобальной системы координат к стандартным системам первой и второй камер осуществляется

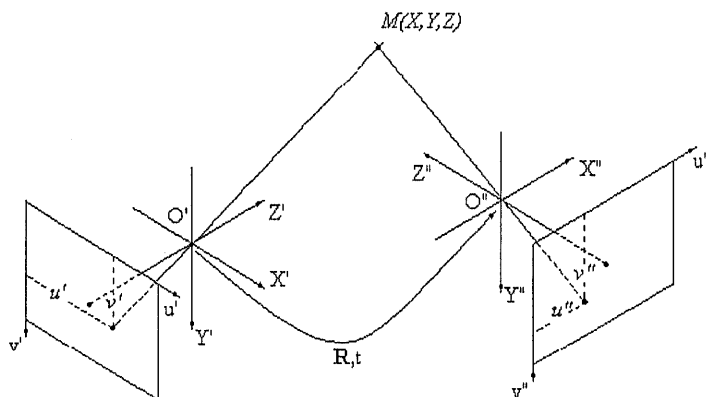


Рис. 6.4. Система двух произвольно ориентированных камер

с помощью преобразований  $M' = R'M + t'$  и  $M'' = R''M + t''$  соответственно. Учитывая это, легко показать, что связь между векторами  $M'$  и  $M''$  задается соотношением

$$M'' = RM' + t, \quad (6.8)$$

где  $R = R''R'^T$  — ортогональная матрица, описывающая ориентацию системы координат второй камеры относительно первой, а  $t = -R''R'^T t' + t''$  — вектор трансляции, определяющий положение оптического центра второй камеры в системе координат первой. Матрицу  $R$  и вектор  $t$  принято называть *внешними параметрами* системы регистрации.

Из выражения (6.8) с помощью (6.2), можно получить соотношение, связывающее координаты сопряженных точек (в координатах фотоприемника):

$$Z''A_2^{-1}v'' = Z'RA_1^{-1}v' + t, \quad (6.9)$$

где (см. (6.2))

$$v' = A_1M' / Z', \quad v'' = A_2M'' / Z''. \quad (6.10)$$

(Предполагаем, что регистрация может выполняться двумя различными камерами, внутренние параметры которых определяются матрицами  $A_1$  и  $A_2$ .)

Уравнения (6.8) и (6.2) позволяют оценить трехмерные координаты точки  $M$  в системе координат любой из камер, если известны внешние параметры системы камер и удастся измерить координаты изображений этой точки в плоскостях изображения камер (т.е. оценить векторы  $v'$  и  $v''$ ). Поскольку компоненты векторов  $v'$  и  $v''$  могут содержать ошибки, реально соотношение (6.9) принимает вид

$$Z''A_2^{-1}v'' - Z'RA_1^{-1}v' - t = e,$$

где  $e$  — вектор невязки, обусловленный наличием ошибок измерений.

Учитывая это, для оценивания неизвестных  $Z'$  и  $Z''$  можно воспользоваться методом наименьших квадратов (МНК), суть которого состоит в том, чтобы найти такие оценки  $Z'$  и  $Z''$ , которые бы минимизировали сумму квадратов компонентов (норму) вектора невязки:  $J_1(Z', Z'') = e^T e$ . Приравнявая частные производные  $J_1$  по  $Z'$  и  $Z''$  нулю (условие достижения экстре-

муму), получаем систему, состоящую из двух скалярных уравнений:

$$\left. \begin{aligned} Z' \mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{A}_1^{-1} \mathbf{v}' - Z' \mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{R}^T \mathbf{A}_2^{-1} \mathbf{v}'' + \mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{R}^T \mathbf{t} &= 0 \\ -Z' \mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{R}^T \mathbf{A}_2^{-1} \mathbf{v}'' + Z'' \mathbf{v}''^T \mathbf{A}_2^{-T} \mathbf{A}_2^{-1} \mathbf{v}'' - \mathbf{v}''^T \mathbf{A}_2^{-T} \mathbf{t} &= 0 \end{aligned} \right\},$$

разрешая которую относительно  $Z'$  и  $Z''$ , находим

$$\begin{bmatrix} Z' \\ Z'' \end{bmatrix} = \begin{bmatrix} \mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{A}_1^{-1} \mathbf{v}' & -\mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{R}^T \mathbf{A}_2^{-1} \mathbf{v}'' \\ -\mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{R}^T \mathbf{A}_2^{-1} \mathbf{v}'' & \mathbf{v}''^T \mathbf{A}_2^{-T} \mathbf{A}_2^{-1} \mathbf{v}'' \end{bmatrix}^{-1} \begin{bmatrix} -\mathbf{v}'^T \mathbf{A}_1^{-T} \mathbf{R}^T \\ \mathbf{v}''^T \mathbf{A}_2^{-T} \end{bmatrix} \mathbf{t}. \quad (6.11)$$

Теперь, используя (6.2), можно вычислить вектор трехмерных координат точки  $M$  в системе любой из камер:

$$\mathbf{M}' = Z' \mathbf{A}_1^{-1} \mathbf{v}', \quad \mathbf{M}'' = Z'' \mathbf{A}_2^{-1} \mathbf{v}''. \quad (6.12)$$

Из всего изложенного выше следует, что для оценивания трехмерных координат некоторой точки по стереопаре необходимо: а) знать внутренние параметры камер (задача калибровки), б) знать параметры взаимного расположения камер (задача взаимного ориентирования), в) найти и определить на изображениях координаты соответствующих данной точке сопряженных точек (задача поиска сопряженных точек).

#### 6.4. КАЛИБРОВКА КАМЕРЫ

Заметим сразу, что, поскольку фокусное расстояние  $f$  и масштабные коэффициенты  $w$  и  $h$  входят в матрицу  $\mathbf{A}$  только в комбинациях  $f/w$  и  $f/h$ , разделить все три этих параметра невозможно. Введем новые параметры  $\alpha_u = f/w$  и  $\alpha_v = f/h$ , которые и будут подлежать оцениванию.

Пусть камера регистрирует сцену, содержащую  $N$  опорных точек, и трехмерные координаты точек известны в глобальной системе координат. Задача состоит в том, чтобы по трехмерным координатам опорных точек  $(X_i, Y_i, Z_i)$  и координатам их проеций в плоскости изображений камеры  $(u_i, v_i)$ ,  $i = \overline{1, N}$  оценить элементы матрицы  $\mathbf{A}$ .

Если положение камеры относительно этой системы известно [т.е. известны матрица  $\mathbf{R}$  и вектор  $\mathbf{t}$  в выражении (6.3)], можно

сразу пересчитать координаты опорных точек в стандартную систему камеры и для оценивания внутренних параметров воспользоваться соотношениями (6.1), подставив вместо  $f/w$  и  $f/h$   $\alpha_u$  и  $\alpha_v$  соответственно. Перепишем (6.1) в виде

$$\alpha_u X + u_0 Z = uZ, \quad \alpha_v Y + v_0 Z = vZ.$$

Для  $N$  опорных точек получим две системы, состоящие из  $N$  линейных уравнений каждая, относительно неизвестных  $u_0$ ,  $\alpha_u$  и  $v_0$ ,  $\alpha_v$ :

$$\begin{aligned} \mathbf{B}_u \mathbf{a}_u &= \mathbf{w}_u, \\ \mathbf{B}_v \mathbf{a}_v &= \mathbf{w}_v, \end{aligned} \tag{6.13}$$

$$\text{где } \mathbf{B}_u = \begin{bmatrix} X_1 & Z_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ X_N & Z_N \end{bmatrix}, \quad \mathbf{a}_u = \begin{bmatrix} \alpha_u \\ u_0 \end{bmatrix}, \quad \mathbf{w}_u = \begin{bmatrix} u_1 Z_1 \\ \cdot \\ \cdot \\ u_N Z_N \end{bmatrix},$$

$$\mathbf{B}_v = \begin{bmatrix} Y_1 & Z_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ Y_N & Z_N \end{bmatrix}, \quad \mathbf{a}_v = \begin{bmatrix} \alpha_v \\ v_0 \end{bmatrix}, \quad \mathbf{w}_v = \begin{bmatrix} v_1 Z_1 \\ \cdot \\ \cdot \\ v_N Z_N \end{bmatrix}.$$

Очевидно, достаточно двух опорных точек, чтобы получить оценки внутренних параметров камеры:

$$\begin{aligned} \alpha_u &= \frac{Z_1 Z_2 (u_1 - u_2)}{X_1 Z_2 - X_2 Z_1}, \quad \alpha_v = \frac{Z_1 Z_2 (v_1 - v_2)}{Y_1 Z_2 - Y_2 Z_1}, \\ u_0 &= \frac{X_2 Z_1 u_1 - X_1 Z_2 u_2}{X_2 Z_1 - X_1 Z_2}, \quad v_0 = \frac{Y_2 Z_1 v_1 - Y_1 Z_2 v_2}{Y_2 Z_1 - Y_1 Z_2}. \end{aligned}$$

Обратим внимание на необходимость выполнения условий  $X_1 Z_2 - X_2 Z_1 \neq 0$  и  $Y_1 Z_2 - Y_2 Z_1 \neq 0$ , которые формально выражают требование, чтобы в плоскости изображения камеры проекции опорных точек не лежали на прямых, параллельных осям координат.

Для увеличения точности оценок внутренних параметров следует использовать большее количество опорных точек. Системы (6.13) в этом случае становятся переопределенными, и для их

решения можно применить МНК. Минимизируя норму вектора невязки  $\mathbf{e} = \mathbf{B}\mathbf{a} - \mathbf{w}$ , получаем оценки внутренних параметров

$$\mathbf{a}_u = (\mathbf{B}_u^T \mathbf{B}_u)^{-1} \mathbf{B}_u^T \mathbf{w}_u, \quad \mathbf{a}_v = (\mathbf{B}_v^T \mathbf{B}_v)^{-1} \mathbf{B}_v^T \mathbf{w}_v.$$

Рассмотрим теперь задачу калибровки в более общей постановке. Пусть неизвестны не только матрица  $\mathbf{A}$ , но и положение камеры относительно глобальной системы координат (т.е. матрица  $\mathbf{R}$  и вектор  $\mathbf{t}$  в выражении (6.3)). Такая постановка обусловлена тем, что технически довольно сложно выполнить точные измерения положения камеры и особенно ее ориентации относительно произвольной системы координат. Используя (6.2) и (6.3) и вводя вектор  $\tilde{\mathbf{M}} = (X, Y, Z, 1)$ , связь между глобальными координатами опорной точки  $M$  и координатами ее изображения можно представить в виде

$$\mathbf{Z}'\mathbf{v}' = \mathbf{A}\mathbf{R}\mathbf{M} + \mathbf{t} = \begin{bmatrix} \alpha_u \mathbf{r}_1^T + u_0 \mathbf{r}_3^T & \alpha_u t_x + u_0 t_z \\ \alpha_v \mathbf{r}_2^T + v_0 \mathbf{r}_3^T & \alpha_v t_y + v_0 t_z \\ \mathbf{r}_3^T & t_z \end{bmatrix} \tilde{\mathbf{M}} = \mathbf{C}\tilde{\mathbf{M}}. \quad (6.14)$$

Матрица  $\mathbf{C}$  имеет размер  $3 \times 4$  и называется *калибровочной*. Обозначим

$$\begin{aligned} \mathbf{c}_1^T &= (c_{11}, c_{12}, c_{13})^T = \alpha_u \mathbf{r}_1^T + u_0 \mathbf{r}_3^T, \quad c_{14} = \alpha_u t_x + u_0 t_z, \\ \mathbf{c}_2^T &= (c_{21}, c_{22}, c_{23})^T = \alpha_v \mathbf{r}_2^T + v_0 \mathbf{r}_3^T, \quad c_{24} = \alpha_v t_y + v_0 t_z, \\ \mathbf{c}_3^T &= (c_{31}, c_{32}, c_{33})^T = \mathbf{r}_3^T, \quad c_{34} = t_z. \end{aligned} \quad (6.15)$$

Отметим два важных свойства этой матрицы, которые легко получить из (6.15), учитывая (6.4) и (6.5):

$$\mathbf{c}_3^T \mathbf{c}_3 = 1, \quad (\mathbf{c}_1 \times \mathbf{c}_3)^T (\mathbf{c}_2 \times \mathbf{c}_3) = 0. \quad (6.16)$$

Оказывается, что если элементы калибровочной матрицы  $c_{ij}$  известны, через них можно вычислить матрицы  $\mathbf{A}$ ,  $\mathbf{R}$  и вектор  $\mathbf{t}$ . Действительно, учитывая (6.4), легко показать, что

$$\begin{aligned} u_0 &= \mathbf{c}_1^T \mathbf{c}_3, \quad v_0 = \mathbf{c}_2^T \mathbf{c}_3, \\ \alpha_u &= \pm |\mathbf{c}_1 \times \mathbf{c}_2|, \\ \alpha_v &= \pm |\mathbf{c}_2 \times \mathbf{c}_3|. \end{aligned} \quad (6.17)$$

(Выбор знака зависит от того, является ли оптическая система камеры оборачивающейся или нет.)



Кроме того, непосредственно из (6.14) следует, что

$$\begin{aligned} t_z &= c_{34} \,, \\ \mathbf{r}_3^T &= \mathbf{c}_3^T \,, \\ t_x &= (c_{14} - u_0 t_z) / \alpha_u \,, \\ t_y &= (c_{24} - v_0 t_z) / \alpha_v \,, \\ \mathbf{r}_1^T &= (\mathbf{c}_1^T - u_0 \mathbf{c}_3^T) / \alpha_u \,, \\ \mathbf{r}_2^T &= (\mathbf{c}_2^T - v_0 \mathbf{c}_3^T) / \alpha_v \,. \end{aligned} \tag{6.18}$$

Задача калибровки, следовательно, сводится к предварительному оцениванию элементов матрицы  $C$ .

Рассмотрим сначала линейный метод оценивания матрицы  $\mathbf{C}$ . Запишем матричное уравнение (6.14) как систему трех обычных уравнений:

$$\begin{aligned} Z'u' &= c_{11}X + c_{12}Y + c_{13}Z + c_{14}, \\ Z'v' &= c_{21}X + c_{22}Y + c_{23}Z + c_{24}, \\ Z' &= c_{31}X + c_{32}Y + c_{33}Z + c_{34} \end{aligned}$$

или, подставляя в два первых уравнения значение  $Z'$  из третьего:

$$c_{11}X + c_{12}Y + c_{13}Z + c_{14} - c_{31}Xu' - c_{32}Yu' - c_{33}Zu' - c_{34}u' = 0,$$

$$c_{21}X + c_{22}Y + c_{23}Z + c_{24} - c_{31}Xv' - c_{32}Yv' - c_{33}Zv' - c_{34}v' = 0.$$

Зная координаты  $(X_i, Y_i, Z_i)$   $N$  опорных точек в трехмерном пространстве и координаты их проекций  $(u_i, v_i)$  в плоскости изображения камеры, получим однородную систему из  $2N$  линейных уравнений относительно 12 неизвестных элементов калибровочной матрицы  $c_{ij}$ :

$$\left. \begin{aligned} c_{11}X_1 + c_{12}Y_1 + c_{13}Z_1 + c_{14} - c_{31}X_1\mu'_1 - c_{32}Y_1\mu'_1 - c_{33}Z_1\mu'_1 - c_{34}\mu'_1 &= 0 \\ c_{21}X_1 + c_{22}Y_1 + c_{23}Z_1 + c_{24} - c_{31}X_1\nu'_1 - c_{32}Y_1\nu'_1 - c_{33}Z_1\nu'_1 - c_{34}\nu'_1 &= 0 \\ . &. . . . . \\ . &. . . . . \\ c_{11}X_N + c_{12}Y_N + c_{13}Z_N + c_{14} - c_{31}X_N\mu'_N - c_{32}Y_N\mu'_N - c_{33}Z_N\mu'_N - c_{34}\mu'_N &= 0 \\ c_{21}X_N + c_{22}Y_N + c_{23}Z_N + c_{24} - c_{31}X_N\nu'_N - c_{32}Y_N\nu'_N - c_{33}Z_N\nu'_N - c_{34}\nu'_N &= 0. \end{aligned} \right\} (6.19)$$

Представим эту систему в матрично-векторном виде:

$$Lc = 0, \quad (6.20)$$

где

$$L = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u'_1 X_1 & -u'_1 Y_1 & -u'_1 Z_1 & -u'_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v'_1 X_1 & -v'_1 Y_1 & -v'_1 Z_1 & -v'_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u'_N X_N & -u'_N Y_N & -u'_N Z_N & -u'_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v'_N X_N & -v'_N Y_N & -v'_N Z_N & -v'_N \end{bmatrix}$$

$$c = (c_{11}, c_{12}, c_{13}, c_{14}, c_{21}, c_{22}, c_{23}, c_{24}, c_{31}, c_{32}, c_{33}, c_{34})^T, \quad 0 = \underbrace{(0, 0, \dots, 0, 0)^T}_{2N}.$$

Сначала рассмотрим некоторые общие особенности этой системы. Очевидно, что одним из решений этой системы является тривиальное  $c = 0$ , которое не имеет физического смысла. Известно [6.2, с.153], что если однородная линейная система имеет хотя бы одно ненулевое решение, то она имеет бесконечное множество решений, причем если  $\hat{c}$  — решение, то и  $a\hat{c}$ , где  $a$  — произвольное число, тоже является решением. Здесь необходимо различать два случая.

Первый — когда ранг матрицы  $L$  на единицу меньше размера вектора  $c$ . Тогда существует только одно (с точностью до произвольного скалярного множителя) решение. Именно этот случай и представляет практический интерес. Для реализации этого условия необходимо (но недостаточно), чтобы количество уравнений в (6.20) было не менее одиннадцати, следовательно, количество опорных точек должно быть не менее шести. Ограничить множество решений можно, воспользовавшись первым из условий (6.16). Действительно, определив некоторое решение  $\hat{c}$ , в качестве оценки компонентов калибровочной матрицы выберем  $c = a\hat{c}$  такое, чтобы  $(a\hat{c}_{31})^2 + (a\hat{c}_{32})^2 + (a\hat{c}_{33})^2 = 1$ . Такая нормировка определяет калибровочную матрицу с точностью до знака. Выбрать правильный знак матрицы можно, например, зная, с какой стороны от плоскости  $OXY$  глобальной системы координат находится камера и учитывая первое из соотношений (6.18). Знак  $c_{34}$  должен совпадать со знаком компонента  $t_z$  вектора трансляции.

Второй случай реализуется, если ранг  $L$  меньше размера вектора  $c$  на два и более. Здесь может существовать множество раз-

личных решений системы, среди которых сделать правильный выбор без привлечения дополнительных данных невозможно. Показано [6.3], что такая ситуация возникает, в частности, если все опорные точки лежат в одной плоскости. Чтобы избежать ее, в качестве тестового объекта часто используют объект, приведенный на рис. 6.5.

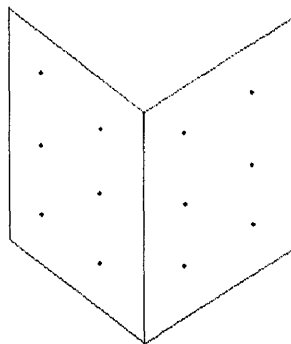


Рис. 6.5. Калибровочный объект

Рассмотрим теперь непосредственно метод решения системы (6.20). Обычно, чтобы уменьшить влияние ошибок измерений трехмерных координат опорных точек и координат их изображений в камере, используют больше, чем шесть опорных точек. Тогда система (6.20) становится переопределенной. Кроме того, как и в разделе 6.2, из-за ошибок в измерениях координат реально система (6.20) принимает вид

$$\mathbf{Lc} = \mathbf{e},$$

где  $\mathbf{e}$ , как и в п. 6.2, — неизвестный вектор невязки, обусловленный наличием ошибок измерений.

В этой ситуации можно снова воспользоваться МНК, согласно которому за оценку вектора  $\mathbf{c}$  следует принять такой, который минимизирует значение функционала  $J_2 = \mathbf{e}^T \mathbf{e} = \mathbf{c}^T \mathbf{L}^T \mathbf{L} \mathbf{c}$  при условии  $\mathbf{c}_3^T \mathbf{c}_3 = 1$  [см. (6.16)]. Так как функционал  $J_2$  представляет собой квадратичную форму с неотрицательно определенной симметричной матрицей  $\mathbf{L}^T \mathbf{L}$ , то минимума по  $\mathbf{c}$  при условии  $\mathbf{c}^T \mathbf{c} \neq 0$  он достигает, если  $\mathbf{c}$  — собственный вектор матрицы  $\mathbf{L}^T \mathbf{L}$ , соответствующий ее минимальному собственному числу  $\lambda_{\min}$  [6.4, с.64]. Заметим, что  $J_{2\min} = \lambda_{\min}$ . (Собственным вектором матрицы  $\mathbf{A}$  называется вектор  $\mathbf{b}$ , который удовлетворяет уравнению  $\mathbf{A}\mathbf{b} = \lambda\mathbf{b}$ , где  $\lambda$  — скалярный множитель, называемый собственным числом матрицы  $\mathbf{A}$ .)

Другой подход к оцениванию матрицы  $\mathbf{C}$  основан на минимизации расстояний между измеренными проекциями опорных точек и вычисленными в соответствии с (6.14). Определим величину

$$E = \sum_{i=1}^N \left\{ \left( \frac{c_{11}X_i + c_{12}Y_i + c_{13}Z_i + c_{14}}{c_{31}X_i + c_{32}Y_i + c_{33}Z_i + c_{34}} - u'_i \right)^2 + \right. \\ \left. + \left( \frac{c_{21}X_i + c_{22}Y_i + c_{23}Z_i + c_{24}}{c_{31}X_i + c_{32}Y_i + c_{33}Z_i + c_{34}} - v'_i \right)^2 \right\},$$

которая представляет сумму квадратов этих расстояний, и минимизируем ее по  $\mathbf{c}$  с учетом ограничений (6.16). Решить эту задачу можно посредством методов условной минимизации [6.5, гл. VI], изложение которых выходит за рамки данного учебника. Следует отметить только, что такой подход обычно дает результаты оценивания, более устойчивые к ошибкам измерений, чем рассмотренный выше линейный.

### 6.5. ВЗАИМНОЕ ОРИЕНТИРОВАНИЕ

В предыдущем разделе показано, как с помощью тестовой сцены не только оценить внутренние параметры камеры, но и определить ее положение относительно системы координат, в которой задана тестовая сцена. Последнее позволяет нам, откалибровав две камеры по одной тестовой сцене, определить их взаимное положение [см. (6.8)]. Однако если внутренние параметры камер определены их конструкцией и могут считаться постоянными, то взаимное положение камер во многих практических случаях может изменяться в процессе регистрации сцены. В связи с этим возникает задача взаимного ориентирования, не предполагающая наличия тестовой сцены.

Рассмотрим снова пару камер, внутренние параметры которых известны, но неизвестны внешние параметры (матрица  $\mathbf{R}$  и вектор  $\mathbf{t}$ ). Умножив обе части выражения (6.8) слева сначала векторно на  $\mathbf{t}$ , а затем скалярно на  $\mathbf{M}''$ , получим

$$\mathbf{M}''^T (\mathbf{t} \times \mathbf{R}\mathbf{M}') = 0.$$

Это соотношение формально выражает тот факт, что векторы  $\mathbf{M}'$ ,  $\mathbf{M}''$  и  $\mathbf{t}$  лежат в одной плоскости, проходящей через три точки: оптические центры камер  $O'$  и  $O''$  и точку наблюдения  $M$ . Выражая  $\mathbf{M}$  через  $\mathbf{v}$  из (6.10), получим:

$$\mathbf{v}''^T \mathbf{A}_2^{-T} (\mathbf{t} \times \mathbf{R}\mathbf{A}_1^{-1} \mathbf{v}') = 0 \quad (6.21)$$

или, учитывая свойства смешанного произведения векторов,

$$(\mathbf{v}^{*T} \mathbf{A}_2^{-T} \times \mathbf{R} \mathbf{A}_1^{-1} \mathbf{v}')^T \mathbf{t} = 0. \quad (6.22)$$

Эквивалентные соотношения (6.21) и (6.22) являются основой для оценивания матрицы  $\mathbf{R}$  и вектора  $\mathbf{t}$ . Предположим, что известны координаты  $n$  пар сопряженных точек и соответственно  $n$  пар векторов  $\mathbf{v}'_i$  и  $\mathbf{v}''_i$   $i = 1, 2, \dots, n$ .

Рассмотрим метод оценивания  $\mathbf{R}$  и  $\mathbf{t}$ , использующий (6.22). Так как это соотношение справедливо для любой пары сопряженных точек, мы имеем систему из  $n$  уравнений относительно неизвестных  $\mathbf{R}$  и  $\mathbf{t}$ , которую можно представить в виде

$$\mathbf{B}_n \mathbf{t} = \mathbf{0}, \text{ где } \mathbf{B}_n = \begin{bmatrix} (\mathbf{v}_1^{*T} \mathbf{A}_2^{-T} \times \mathbf{R} \mathbf{A}_1^{-1} \mathbf{v}'_1)^T \\ \dots\dots\dots \\ (\mathbf{v}_n^{*T} \mathbf{A}_2^{-T} \times \mathbf{R} \mathbf{A}_1^{-1} \mathbf{v}'_n)^T \end{bmatrix}. \quad (6.23)$$

Система (6.23) является однородной, линейной по  $\mathbf{t}$ . Это означает, что вектор трансляции можно оценить только с точностью до постоянного множителя. Вводя условие нормировки  $\|\mathbf{t}\|^2 = 1$ , количество возможных решений можно ограничить двумя, отличающимися знаком. Вопрос о выборе знака будет рассмотрен позже. Система (6.23) содержит пять неизвестных, так как матрица  $\mathbf{R}$  в силу условий нормировки и ортогональности зависит от трех параметров, а вектор  $\mathbf{t}$  с учетом введенной нормировки — от двух. Поэтому количество уравнений в системе, следовательно, и число пар известных сопряженных точек  $n$  должно быть не менее пяти.

Поскольку на практике в матрицу  $\mathbf{B}_n$  входят не точные значения координат сопряженных точек, а результаты их измерений, которые могут содержать ошибки, реально система (6.23) имеет ненулевую правую часть, т.е.

$$\mathbf{B}_n \mathbf{t} = \mathbf{e},$$

где  $\mathbf{e}$ , как и в п. 6.2, — вектор невязки, обусловленный наличием ошибок измерений.

Согласно МНК в качестве оценок матрицы вращения и вектора трансляции следует выбрать такие  $\mathbf{R}$  и  $\mathbf{t}$ , которые минимизируют значения функционала  $J_2(\mathbf{R}, \mathbf{t}) = \mathbf{e}^T \mathbf{e} = \mathbf{t}^T \mathbf{B}_n^T \mathbf{B}_n \mathbf{t}$ . Как упоминалось ранее, при условии  $\|\mathbf{t}\|^2 = 1$  квадратичная форма  $\mathbf{t}^T \mathbf{B}_n^T \mathbf{B}_n \mathbf{t}$  достигает минимума  $J_2 = \lambda_{\min}$  по  $\mathbf{t}$  ( $\lambda_{\min}$  — минималь-

ное собственное число матрицы  $\mathbf{B}_n^T \mathbf{B}_n$ ), если  $\mathbf{t}$  — собственный вектор матрицы, соответствующий  $\lambda_{\min}$ . Поэтому процедуру оценивания  $\mathbf{R}$  и  $\mathbf{t}$  можно разбить на два этапа. На первом находится матрица  $\mathbf{R}$ , минимизирующая  $\lambda_{\min}$ . На втором оценивается собственный вектор матрицы  $\mathbf{B}_n^T \mathbf{B}_n$ , соответствующий  $\lambda_{\min}$ . Существует множество алгоритмов и их программных реализаций для вычисления собственных векторов, поэтому второй этап не вызывает трудностей.

Значительно более сложная задача — задача оценивания матрицы  $\mathbf{R}$ . Один из возможных алгоритмов состоит в следующем [6.6]. Известно [6.1, п.14.10], что матрица  $\mathbf{R}$  может быть представлена в виде

$$\mathbf{R} = \mathbf{R}_x(\omega_x) \mathbf{R}_y(\omega_y) \mathbf{R}_z(\omega_z),$$

где

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega_x & -\sin \omega_x \\ 0 & \sin \omega_x & \cos \omega_x \end{bmatrix}, \quad \mathbf{R}_y = \begin{bmatrix} \cos \omega_y & 0 & \sin \omega_y \\ 0 & 1 & 0 \\ -\sin \omega_y & 0 & \cos \omega_y \end{bmatrix},$$

$$\mathbf{R}_z = \begin{bmatrix} \cos \omega_z & -\sin \omega_z & 0 \\ \sin \omega_z & \cos \omega_z & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Углы  $\omega_x$ ,  $\omega_y$  и  $\omega_z$  и есть те три неизвестных параметра, от которых зависит матрица  $\mathbf{R}$ . На практике всегда известен диапазон, в котором они могут лежать. Выполняя в этом диапазоне полный перебор по всем углам с достаточно грубым шагом (например,  $1^\circ$ ), можно приблизиться к значениям, удовлетворяющим требованиям минимизации функционала  $J_2$  по  $\mathbf{R}$ . Затем в окрестности этих значений для уточнения положения минимума можно воспользоваться одним из известных методов минимизации [6.5, гл.V] (например, наискорейшего спуска, Ньютона, Маркуардта).

Наконец, получив оценки  $\mathbf{R}$  и  $\mathbf{t}$ , можно, используя (6.11), оценить и  $Z$  — координаты наблюдаемых точек. Из способа задания систем координат (см. рис. 6.4) следует, что  $Z'$  и  $Z''$  должны быть положительными. Этим условием и определяется выбор правильного знака вектора трансляции  $\mathbf{t}$ .

В заключение необходимо сказать, что развитием темы взаимного ориентирования является задача *самокалибровки* системы камер, цель которой – оценивание как внутренних, так и внешних параметров. Не останавливаясь на этой задаче подробно, отметим только, что в системе, состоящей из двух, даже одинаковых, камер, данных для самокалибровки недостаточно. Добавление третьей камеры с теми же внутренними параметрами делает самокалибровку возможной. Подробное исследование этого вопроса можно найти в работе [6.7].

Отметим, что задача определения взаимного положения камер может иметь и другую трактовку. Предположим, что движущаяся камера непрерывно регистрирует некоторую сцену. Тогда, анализируя последовательность изображений и решая эту задачу, можно определить характер движения камеры в пространстве.

## 6.6. ПОИСК СОПРЯЖЕННЫХ ТОЧЕК

Центральная проблема компьютерного стереозрения – это поиск сопряженных точек. Задача автоматизированного поиска сопряженных точек состоит в следующем: на одном из изображений выбрана точка  $m'$ , являющаяся проекцией некоторой точки  $M$  трехмерного пространства, необходимо на втором изображении найти точку  $m''$  – проекцию той же точки. Достаточно очевидно, что фактически при решении этой задачи речь идет об отождествлении не отдельных точек, а фрагментов изображений, лежащих в окрестностях этих точек. Здесь возникает два вопроса. Первый – если на одном изображении указана точка, где на втором изображении искать сопряженную? Второй – что является критерием «похожести» фрагментов?

Оказывается, на первый вопрос можно дать достаточно строгий ответ. Обратимся к соотношению (6.21). Известно [6.8, приложение], что векторное произведение  $\mathbf{a} \times \mathbf{b}$  можно представить как  $[\mathbf{a}]_{\times} \mathbf{b}$ , где матрица  $[\mathbf{a}]_{\times}$  имеет вид

$$[\mathbf{a}]_{\times} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}.$$

Тогда (6.21) будет иметь вид

$$\mathbf{v}^{\prime T} \mathbf{F} \mathbf{v}' = 0 \quad (6.24)$$

или

$$\mathbf{v}'^T \mathbf{F}^T \mathbf{v}'' = 0,$$

$$\text{где } \mathbf{F} = \mathbf{A}_2^{-T} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{A}_1^{-1}. \quad (6.25)$$

С другой стороны, любая прямая линия на плоскости может быть задана уравнением  $ax + by + c = 0$  или в векторном виде,  $\mathbf{a}^T \mathbf{m} = 0$ , где  $\mathbf{m} = (x, y, 1)$ ,  $\mathbf{a} = (a, b, c)$ . Сравнивая уравнение прямой с (6.24) и с (6.25), приходим к выводу, что (6.24) задает в плоскости изображения первой камеры прямую

$$\mathbf{a}'^T \mathbf{v}' = 0 \quad (6.26)$$

с вектором коэффициентов  $\mathbf{a}' = \mathbf{F}^T \mathbf{v}''$ , а в плоскости изображения второй камеры — прямую

$$\mathbf{a}''^T \mathbf{v}'' = 0 \quad (6.27)$$

с вектором коэффициентов  $\mathbf{a}'' = \mathbf{F} \mathbf{v}'$ .

Это значит, что если в плоскости изображения первой камеры указана точка с вектором внутренних координат  $\mathbf{v}'$ , то сопряженная точка в плоскости изображения второй камеры может лежать только на прямой, заданной уравнением (6.26). Аналогичное правило справедливо и для точек, заданных в плоскости изображения второй камеры.

Этот результат имеет наглядную геометрическую иллюстрацию (рис. 6.6).

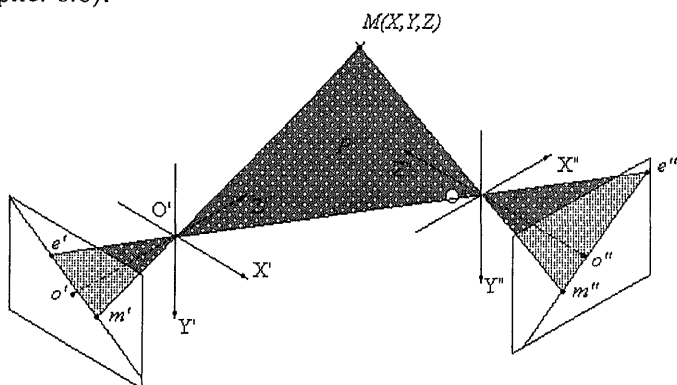


Рис. 6.6. Эпиполярные линии

Оптические центры камер  $O'$  и  $O''$  и наблюдаемая точка  $M$  образуют плоскость  $P$ , которая пересекает плоскости изображе-



ния камер по прямым  $e'm'$  и  $e''m''$ , где точка  $e'$  является изображением оптического центра  $O''$  второй камеры в плоскости первой, а  $e''$  — изображением центра  $O'$  в плоскости второй. С другой стороны, если задана точка  $m'$ , то точка  $M$  трехмерного пространства может лежать только на прямой, проходящей через  $m'$  и оптический центр  $O'$ . Следовательно, ее проекция в плоскости изображения второй камеры должна лежать в плоскости  $P$ , лежащей на этой прямой и оптическом центре  $C''$ , а значит, на линии пересечения плоскости  $P$  и плоскости изображения второй камеры. Линии пересечения плоскости  $P$  с плоскостями изображений камер называются *эпиполярными линиями*.

Практическая ценность полученного результата заключается в том, что его использование при поиске сопряженных точек позволяет существенно сократить размер зоны поиска, выполняя поиск не на всем изображении, а только вдоль эпиполярной линии, и тем самым снизить вероятность ложной идентификации фрагментов.

В отличие от рассмотренных ранее вопросов, сопоставление окрестностей сопряженных точек не поддается строгой формализации, поскольку в его основе лежит проблема идентификации по изображениям фрагментов реального трехмерного мира, которые с трудом поддаются адекватному формальному описанию. Возникающие при съемке и зависящие от ракурса проективные и яркостные искажения приводят к тому, что в изображениях одних и тех же участков сцены, снятых под разными ракурсами, могут появиться значительные отличия. Принципиально важно то, что эти отличия зависят не только от геометрии съемки, но и от геометрических и физических характеристик самой сцены. Расположение источника света по отношению к поверхности сцены влияет на распределение освещенности. Положение элементов сцены, их отражательные и рассеивающие свойства определяют количество энергии, попадающей в объективы камер, а следовательно, и локальные отличия в яркости сопряженных фрагментов изображений.

На рис. 6.7 схематично показаны некоторые причины возникновения различий в изображениях. Пунктиром показаны индикатрисы излучения (рассеяния) в точках  $A$  и  $B$  наблюдаемой поверхности. Видно, что интенсивность излучения из точки  $A$  в направлении левой камеры (с оптическим центром  $C'$ ) больше, чем в направлении правой камеры (с оптическим центром  $C''$ ). Точка  $P$ , наоборот, излучает в камеру  $C'$  слабее, чем в камеру  $C''$ . Жирными столбиками показана интенсивность сигнала в изо-

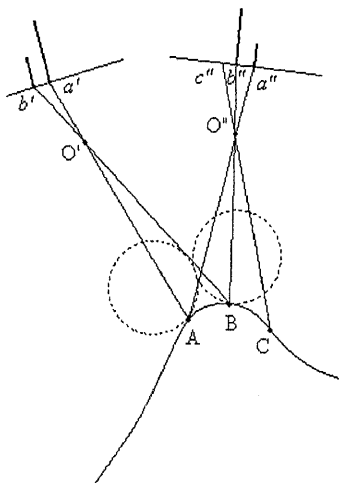


Рис. 6.7. Проективные и яркостные искажения

бражениях этих точек в левой ( $a'$  и  $b'$ ) и правой ( $a''$  и  $b''$ ) камерах. Более того, точка  $C$  вообще не видна в левой камере, поскольку ее загораживает участок поверхности  $BC$ .

Величина различий зависит от разницы в ракурсах съемки. Как правило, чем больше эта разница (в частности, чем больше база), тем менее похожими становятся изображения. Поэтому все методы отождествления окрестностей сопряженных точек в большей или меньшей степени опираются не на формальный подход, а на здравый смысл.

Одним из наиболее распространенных методов идентификации является корреляционный. Детальное описание корреляционного метода приведено в гл.5. При незначи-

тельных отличиях в ракурсах съемки и на достаточно гладких поверхностях от него можно ожидать хороших результатов. На рис. 6.8 ( $a$  и  $b$ ) приведена пара вертолетных снимков района озера Карымское (Камчатка), на которые нанесена прореженная сетка сопряженных точек, найденных корреляционным методом, а на рис. 6.8,  $в$  показана поверхность, «натяннутая» на трехмерные точки, координаты которых вычислены методом, изложенным в п. 6.2. Следует заметить, что построение поверхности, лежащей на заданном множестве трехмерных точек, представляет собой самостоятельную нетривиальную задачу, рассмотрение которой выходит за рамки темы, обсуждаемой в этой главе.

На снимках есть области, где сопряженные точки не удалось найти. Причин этому несколько. Часть сцены, видимая справа на снимке 6.8,  $б$  просто не попала в поле зрения левой камеры. Кроме этого, на снимках есть участки с мало изменяющейся яркостью. На таких участках корреляционный критерий не дает хорошо выраженного максимума, поэтому здесь координаты сопряженных точек определяются с большими погрешностями. Более того, здесь велика вероятность ложных отождествлений. Пространственное представление о таких участках получить без привлечения дополнительной информации невозможно. Если размеры этих участков малы, а по косвенным признакам можно предположить, что поверхность сцены изменяется плавно, мож-

но интерполировать трехмерную структуру участков с их границ внутри, как это сделано для участков  $a', a''$  и  $b', b''$ . Если их размеры велики или сопряженные точки на границах не определены, восстановить трехмерную структуру участка не удастся (участки  $c', c''$  и  $d', d''$ ).

Для того чтобы восстановить мелкую структуру сцены, необходимо, чтобы сопряженные точки лежали достаточно часто. В приведенном примере для построения рельефа найдено более 7000 точек.

Сложность применения корреляционного метода заключается в том, чтобы подобрать такие размеры сопоставляемых фрагментов, при которых отличия в тождественных фрагментах еще невелики (для этого нужно уменьшать размеры), а оценка коэффициента корреляции остается достоверной (для этого размеры надо увеличивать). Но самым серьезным его недостатком является чувствительность к масштабным искажениям, присутствующим в отождествляемых фрагментах. Простейший способ уменьшения масштабных различий в сопряженных фрагментах состоит в использовании предварительной аффинной подстройки изображений. Этот способ, однако, оказывается недейственным при вариациях высот в сцене, сравнимых с расстоянием от сцены до системы камер, поскольку в этом случае появляются значительные локальные масштабные искажения.

Влияние яркостных искажений можно попытаться устранить, отказавшись от сопоставления фрагментов исходных изображений и перейдя к сопоставлению фрагментов контурных изображений, где выделены края, т.е. переходы между областями с постоянной яркостью. На таких переходах, как известно, градиент яркости имеет максимум, а лапласиан обращается в нуль. Это обстоятельство и используется для выделения краев [6.10, п.17.4]. Существенным моментом является то, что контуры на изображениях непрерывных поверхностей должны располагаться в одинаковом порядке, как это видно на рис. 6.9 (изображения точек  $A$ ,  $B$  и  $D$ ). Это ограничение позволяет в некоторой степени ослабить влияние масштабных искажений. Однако и здесь, как и в корреляционном методе, существует проблема ложного отождествления. Связана она, в частности, с тем, что видимая область поверхности зависит от ракурса наблюдения (поэтому, например, на рис. 6.9 изображение точки  $C$  в правой камере отсутствует) и на разных изображениях могут пропадать разные края. Кроме того, сама задача выделения краев только на первый взгляд кажется такой простой. В действительности неизбежное

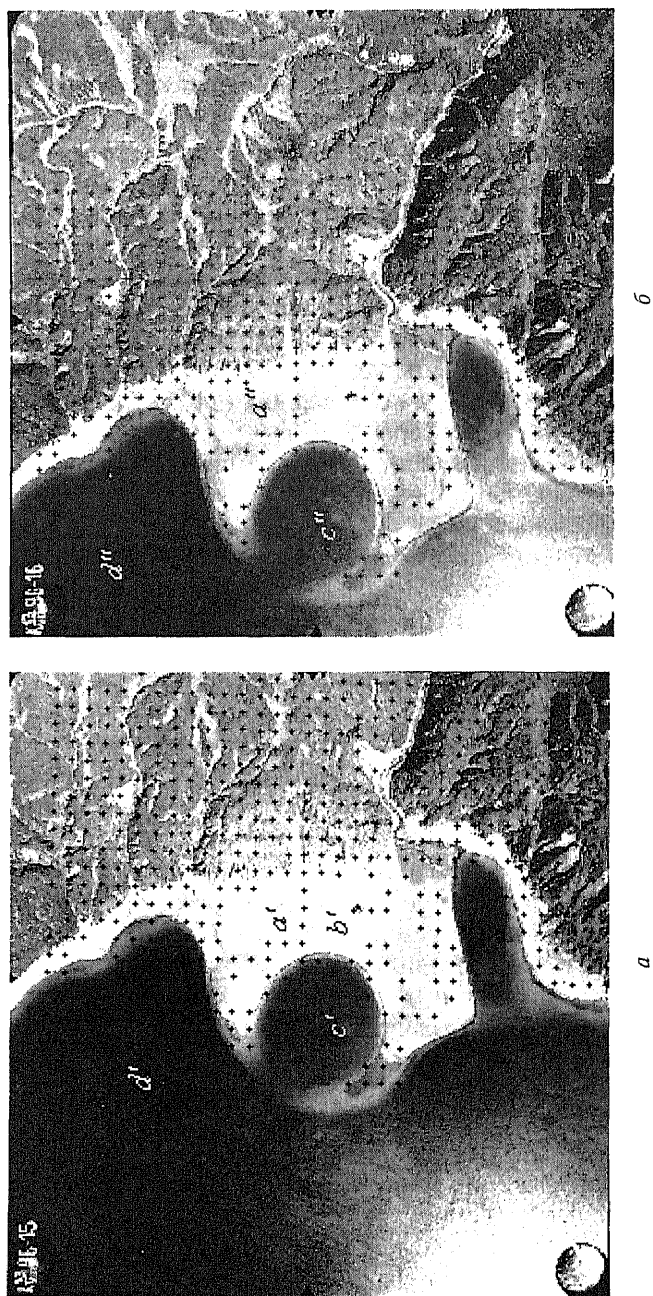
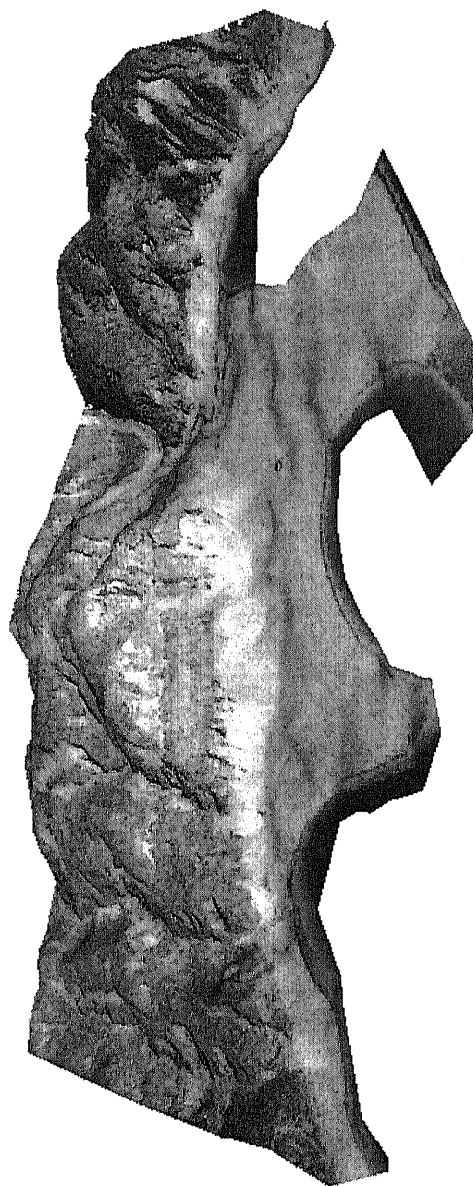


Рис. 6.8. Восстановление поверхности трехмерной сцены по стереопаре (приведено с любезного согласия авторов [6,9])(см. также с. 191)



в

Рис. 6.8. Окончание

присутствие шумов в изображениях может привести к частично-му исчезновению контуров или появлению ложных. Идея поиска сопряженных точек с помощью отождествления краев была особенно популярна у биологов и психологов, исследовавших механизмы зрения [6.11], но, как нам кажется, в таком идеальном виде практического применения не нашла.

Еще один подход к поиску сопряженных точек опирается на предварительный отдельный анализ изображений, позволяющий выделить на них некоторые характерные объекты или особенности. Так на аэроснимках городских территорий можно попытаться выделить перекрестки, отдельные дома или их фрагменты, деревья и другие объекты, а затем выполнить их отождествление между снимками.

По-видимому, наиболее удачных результатов можно ожидать от комбинированного использования всех этих методов и применения итеративной схемы, когда по найденным сопряженным точкам строится оценка поверхности, с учетом которой изображения подвергаются масштабной коррекции и выполняется поиск новых сопряженных точек.

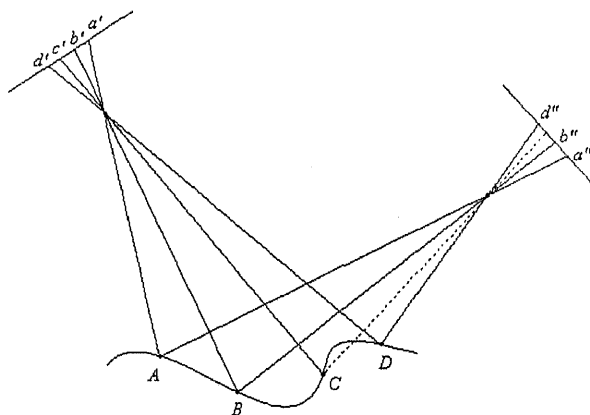


Рис. 6.9. Упорядочение сопряженных точек в случае непрерывных поверхностей

## ВОПРОСЫ К ГЛАВЕ 6

- 6.1. Какой смысл имеет знак при  $f$  в выражении (6.1)?
- 6.2. Какой вид примет выражение (6.3), если вектор трансляции будет задан в глобальной системе координат?

6.3. Обратимся к рис. 6.3. Какой вид будут иметь матрицы  $A_1$ ,  $A_2$ ,  $R$  и вектор  $t$  в выражении (6.9) для ситуации, приведенной на рисунке?

6.4. Пусть в ситуации, показанной на рис. 6.3, векторы  $v' = (x', y, 1)$  и  $v'' = (x'', y, 1)$  внутренних координат проекций точки  $M$  в плоскостях изображений левой и правой камер известны. Получите оценки трехмерных координат точки  $M$  в системах координат правой и левой камер, пользуясь выражениями (6.11) и (6.12). Сравните полученный результат с (6.6) и (6.7). Объясните отличия.

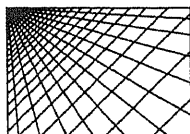
6.5. Почему в предыдущем вопросе векторы  $v' = (x', y, 1)$  и  $v'' = (x'', y, 1)$  имеют одинаковые  $y$  — компоненты?

6.6. Докажите справедливость соотношений (6.17).

6.7. Докажите справедливость соотношений (6.18).

6.8. На рис. 6.6 точки  $e'$  и  $e''$  являются изображениями оптических центров камер. Полагая, что матрицы  $A_1$ ,  $A_2$ ,  $R$  и вектор  $t$  известны, найдите координаты этих точек в плоскостях изображений соответствующих камер.

6.9. Где будут находиться точки  $e'$  и  $e''$ , если камеры расположены, как показано на рис. 6.3?



## ГЛАВА 7

### СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ

Часто изображение представляется для наблюдателя в виде совокупности однородных участков, отличающихся друг от друга теми или иными характеристиками. Количество *различных типов*, или *классов*, участков обычно невелико, вся же картина разделена на непересекающиеся области, каждая из которых заполнена изображением одного из типов. При анализе таких изображений наблюдателем или автоматической системой целью является *определение геометрических областей* и указание для каждой из них *номера типа (класса)*. Иногда такую совокупность сведений об исходном изображении называют его *картой*. Обработка изображения, позволяющая получить карту, называется *сегментацией*.

Количество признаков, которыми могут отличаться друг от друга участки различных типов, очень велико. Встречается большое число задач, где участки имеют неодинаковую среднюю яркость. Иногда при одинаковой средней яркости различаются дисперсии флюктуаций. Часто можно наблюдать картину, на которой различие областей проявляется в неодинаковости корреляционных свойств: медленные, плавные колебания яркости на одних участках сменяются существенно более быстрыми на других. Примером может служить аэрофотоснимок земной поверхности, на котором участки леса чередуются с участками поля или водной поверхности. Картина такого рода показана в конце данной главы на рис. 7.6, а. Во многих случаях участки разных типов различаются не по одной какой-либо характеристике, а по нескольким.

Встречаются изображения, в которых вся картина разбита на области, не отличающиеся друг от друга ни по каким своим характеристикам. Вся информация (карта) в этих случаях пред-



ставляет собой совокупность границ между отдельными участками. Типичным примером такого типа служит изображение кирпичной кладки. Очевидно, что для сегментации таких изображений непригодны методы, основанные на анализе моментно-корреляционных характеристик, о которых говорилось выше.

Большое разнообразие в задаче сегментации вносит обработка цветных изображений. Одним из важнейших признаков для сегментации в этом случае может служить цвет, который дополняет совокупность характеристик, применяемых при обработке черно-белых изображений. К этому же типу проблем относится и сегментация спектрально-зональных изображений. Отметим, что такие задачи значительно сложнее, а число публикаций в этой области сравнительно невелико. По этим причинам в данном учебном пособии сегментация цветных и спектрально-зональных изображений не рассматривается. Вместе с тем подчеркнем, что такие задачи представляют очень большой научный и практический интерес и, вероятно, они составят одно из перспективных направлений развития цифровой обработки изображений.

Методы, используемые при сегментации, еще более разнообразны, чем признаки, по которым различаются отдельные классы. Следует отметить, что единого общепризнанного, эффективного подхода, который бы лежал в основе всех или хотя бы большинства методов, не существует. Поэтому трудно в рамках учебного издания дать полную картину методов. Известны отдельные идеи, на основе которых решаются многие задачи сегментации. Некоторые подходы представляются авторам более перспективными, чем другие. Задача данного пособия — познакомить читателя с распространенными или перспективными идеями, что могло бы послужить импульсом для самостоятельного изучения других методов, а также составить основу для собственных исследований и разработок в данной области.

### 7.1. СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ПОРОГОВОЙ ОБРАБОТКИ

Пороговая обработка, вероятно, самый простой метод сегментации, что привлекает к нему большое внимание специалистов. Метод ориентирован на обработку изображений, отдельные однородные участки которых различаются средней яркостью. Простейшим и вместе с тем часто применяемым видом сегментации является *бинарная сегментация*, когда имеется только два типа однородных участков. При этом преобразование каждой

точки исходного изображения  $x(i_1, i_2)$  в выходное  $y(i_1, i_2)$  выполняется по правилу:

$$y(i_1, i_2) = \begin{cases} y_0, & \text{при } x(i_1, i_2) \leq x_0 \\ y_1, & \text{при } x(i_1, i_2) > x_0, \end{cases} \quad (7.1)$$

где  $x_0$  — единственный параметр обработки, называемый порогом. Уровни выходной яркости  $y_0$  и  $y_1$  могут быть произвольными, они лишь выполняют функции *меток*, при помощи которых осуществляется разметка получаемой карты — отнесение ее точек к классам  $H_0$  или  $H_1$  соответственно. Если образуемый препарат подготавливается для визуального восприятия, то часто их значения соответствуют уровням черного и белого. Если существует более двух классов, то при пороговой обработке должно быть задано семейство порогов, отделяющих яркости различных классов друг от друга.

Центральным вопросом пороговой сегментации является определение порогов, которое должно выполняться автоматически. Применяемые в настоящее время методы автоматического определения порогов подробно описаны в обзоре [7.1]. Разнообразие методов очень велико, однако в основном они базируются на анализе гистограммы исходного изображения.

Пусть  $w(x)$ ,  $0 \leq x \leq 255$  — гистограмма исходного цифрового изображения. Примем, что его диапазон яркостей заключен в пределах от 0 (уровень черного) до 255 (уровень белого). Первоначальная идея гистограммного метода определения порога основывалась на предположении о том, что распределения вероятностей для каждого класса унимодальны (содержат по одному пику), а точки границ, разделяющих участки разных классов на изображении, малочисленны. Этим предположениям должна отвечать гистограмма, которая имеет многомодальный характер. Отдельные моды соответствуют различным классам, а разделяющие их впадины — малочисленным по количеству входящих в них точек граничным областям. Пороги сегментации находятся при этом по положению впадин. Рис. 7.1 иллюстрирует сказанное выше применительно к случаю двух классов. В действительности воспользоваться такими простыми соображениями для выбора порога удастся крайне редко. Дело в том, что реальные гистограммы обычно сильно изрезаны, что иллюстрирует приводимый на рис. 7.2, в результат эксперимента. Это служит первым препятствием для определения точек минимума. Вторым препятствием является то, что границы между однородными участ-

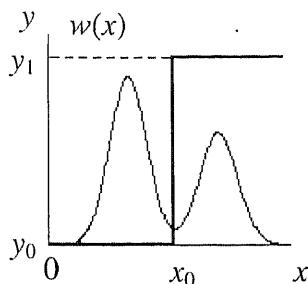


Рис. 7.1. К выбору порога бинарной сегментации

ками на изображении бывают размыты, вследствие чего уровень гистограммы в тех ее частях, которые отображают точки границы, возрастает. Очевидно, это приводит к уменьшению провалов в гистограмме или даже их исчезновению.

Один из эффективных путей преодоления этих трудностей состоит в определении порога на основе так называемого *дискриминантного критерия*. Рассмотрим этот подход применительно к двум классам, поскольку обобщение на случай большего числа классов не составляет принципиальной проблемы. Итак, считаем, что распределение  $w(x)$ ,  $0 \leq x \leq 255$  построено для изображения, содержащего два типа участков, причем существует оптимальная граница  $x_0$ , разделяющая их наилучшим образом в некотором смысле. Для определения оптимального порога строим *дискриминантную функцию*  $\eta(x_0^*)$ ,  $0 \leq x_0^* \leq 255$ , аргумент которой  $x_0^*$  имеет смысл пробного порога. Его значение, максимизирующее функцию  $\eta(x_0^*)$ , является оптимальным порогом  $x_0$ . Рассмотрим построение дискриминантной функции.

Пусть  $x_0^*$  — гипотетическое значение порога, разбивающее распределение  $w(x)$ ,  $0 \leq x \leq 255$  на два класса. При этом обычно не играет большой роли, к какому из классов будут отнесены точки изображения, имеющие яркость  $x_0^*$ , в силу малочисленности граничных точек, разделяющих участки разных классов. Вероятность того, что наугад взятая точка кадра принадлежит классу  $H_0$ , равна

$$p_0 = \sum_{x=0}^{x_0^*} w(x) = p(x_0^*). \quad (7.2)$$

Аналогично вероятность ее принадлежности к классу  $H_1$  определяется формулой

$$p_1 = \sum_{x=x_0^*+1}^{255} w(x), \quad (7.3)$$

причем в силу нормировки распределения вероятностей имеет место равенство

$$p_0 + p_1 = 1.$$

Далее считаем, что участок распределения  $w(x)$ ,  $0 \leq x \leq x_0^*$ , ограниченный точкой  $x_0^*$ , описывает часть изображения, принадлежащую  $H_0$ , а участок  $w(x)$ ,  $x_0^* + 1 \leq x \leq 255$  — принадлежащую  $H_1$ . Это позволяет ввести в рассмотрение два распределения  $w_0(x)$  и  $w_1(x)$ , соответствующих  $H_0$  и  $H_1$ , конструируя их из распределения  $w(x)$  при помощи выражений:

$$w_0(x) = w(x)/p_0, \quad 0 \leq x \leq x_0^*$$

$$w_1(x) = w(x)/p_1, \quad x_0^* + 1 \leq x \leq 255.$$

Здесь делением на вероятности  $p_0$  и  $p_1$  обеспечивается нормировка вводимых условных распределений.

Для образованных таким образом распределений вероятностей могут быть найдены моменты. Выражения для математических ожиданий  $\mu_0$  и  $\mu_1$  имеют вид

$$\mu_0 = \sum_{x=0}^{x_0^*} x w_0(x) = \sum_{x=0}^{x_0^*} x w(x) / p_0, \quad (7.4)$$

$$\begin{aligned} \mu_1 &= \sum_{x=x_0^*+1}^{255} x w_1(x) = \sum_{x=x_0^*+1}^{255} x w(x) / p_1 = \\ &= \left[ \sum_{x=0}^{255} x w(x) - \sum_{x=0}^{x_0^*} x w(x) \right] / p_1 = \\ &= \left[ \mu(255) - \mu(x_0^*) \right] / p_1 = \frac{\mu(255) - \mu(x_0^*)}{1 - p(x_0^*)}, \end{aligned} \quad (7.5)$$

где  $\mu(x_0^*) = \sum_{x=0}^{x_0^*} xw(x)$  — ненормированное математическое ожидание для  $H_0$ ,  $\mu(255) = \sum_{x=0}^{255} xw(x)$  — математическое ожидание для всего кадра.

Аналогично, дисперсия для всего кадра  $\sigma_T^2$  определяется выражением

$$\sigma_T^2 = \sum_{x=0}^{255} [x - \mu(255)]^2 w(x). \quad (7.6)$$

Для построения дискриминантной функции дополнительно вводим еще один энергетический параметр  $\sigma_B^2$ , называемый *межклассовой дисперсией*:

$$\sigma_B^2 = p_0(\mu_0 - \mu(255))^2 + p_1(\mu_1 - \mu(255))^2. \quad (7.7)$$

Безразмерная дискриминантная функция определяется выражением

$$\eta(x_0^*) = \frac{\sigma_B^2(x_0^*)}{\sigma_T^2}. \quad (7.8)$$

Оптимальным, как говорилось выше, считается порог, отвечающий требованию

$$x_0 = \arg \max_{x_0^*} \eta(x_0^*). \quad (7.9)$$

Поясним смысл критерия (7.9). Знаменатель  $\sigma_T^2$  в выражении (7.8) является дисперсией всего кадра и, следовательно, от величины пробного порога  $x_0^*$ , разбивающего изображение на классы, не зависит. Поэтому точка максимума выражения (7.8) совпадает с точкой максимума числителя, т.е. определяется характером зависимости межклассовой дисперсии (7.7) от порога  $x_0^*$ . При его стремлении к нулю вероятность  $p_0$ , как следует из (7.2), также стремится к нулю. Поскольку при этом все изображение относится к классу  $H_1$ , имеет место тенденция  $\mu_1 \rightarrow \mu(255)$ . Следовательно, оба слагаемых в (7.7) становятся равными нулю. Это же наблюдается и при другом крайнем значении порога  $x_0^* = 255$ . В силу неотрицательности величин, входящих в (7.7) и

(7.9), и равенства функции  $\eta(x_0^*)$  нулю на краях области определения, внутри этой области существует максимум, абсцисса которого и принимается за оптимальный порог. Следует отметить качественный характер этих соображений. Более детальные исследования показывают, например, что при обработке некоторых изображений дискриминантная функция  $\eta(x_0^*)$  имеет несколько максимумов даже при наличии на изображении только двух классов. Это, в частности, проявляется, когда суммарные площади участков, занятых классами  $H_0$  и  $H_1$ , существенно различны. Поэтому задача в общем случае несколько усложняется необходимостью определить *абсолютный* максимум функции  $\eta(x_0^*)$ .

С вычислительной точки зрения для выполнения алгоритма необходимо найти для всего изображения математическое ожидание  $\mu(255)$  и дисперсию  $\sigma_T^2$ . Далее при каждом значении  $x_0^*$  определяются вероятности  $p_0$  и  $p_1$  с использованием (7.2) и (7.3) (или условия нормировки), а также математические ожидания классов  $\mu_0$  и  $\mu_1$  при помощи соотношений (7.4), (7.5). Найденные таким образом величины дают возможность определить значение  $\eta(x_0^*)$ .

Объем вычислений можно сократить, если выполнить некоторые преобразования формулы (7.7) для межклассовой дисперсии. Используя формулы (7.2)...(7.5), нетрудно получить соотношение для математических ожиданий:

$$p_0\mu_0 + p_1\mu_1 = \mu(255). \quad (7.10)$$

Раскрывая далее в (7.7) квадраты разностей, после преобразований с использованием соотношения (7.10), можно получить:

$$\sigma_B^2 = p_0p_1(\mu_1 - \mu_0)^2. \quad (7.11)$$

Выражая из (7.10) величину  $\mu_1$  и подставляя ее в (7.11), окончательно находим:

$$\sigma_B^2(x_0^*) = \frac{[\mu(255)p(x_0^*) - \mu(x_0^*)]^2}{p(x_0^*)[1 - p(x_0^*)]}. \quad (7.12)$$

В соотношение (7.12), используемое в качестве рабочего, входят лишь две величины — вероятность  $p(x_0^*)$  и ненормированное

математическое ожидание  $\mu(x_0^*)$ , что существенно уменьшает объем вычислений при автоматическом отыскании оптимального порога.

На рис. 7.2 приведены результаты эксперимента, иллюстрирующие описанный метод автоматической бинарной сегментации. На рис.7.2, *а* показан аэрофотоснимок участка земной поверхности “Поле”, а на рис.7.2, *б* – результат его бинарной сегментации, выполненной на основе автоматического определения порога при помощи дискриминантного метода. Гистограмма распределения исходного изображения показана на рис.7.2, *в*, а дискриминантная функция  $\eta(x_0^*)$ , вычисленная по полученной гистограмме – на рис. 7.2, *г*. Сильная изрезанность гистограммы, порождающая большое количество минимумов, исключает возможность непосредственного определения единственного информационного минимума, разделяющего классы друг от друга. Функция же  $\eta(x_0^*)$  является существенно более гладкой и к тому же в данном случае унимодальной, что делает определение порога весьма простой задачей. Оптимальный порог, при котором получено сегментированное изображение,  $x_0 = 100$ . Результаты показывают, что описанный метод нахождения порога, являясь развитием гистограммного подхода, обладает сильным сглаживающим действием по отношению к изрезанности самой гистограммы.

Коснемся вопроса о пороговой сегментации нестационарных изображений. Если средняя яркость изменяется внутри кадра, то пороги сегментации должны быть также изменяющимися. Часто в этих случаях прибегают к разбиению кадра на отдельные области, в пределах которых изменениями средней яркости можно пренебречь. Это позволяет применять внутри отдельных областей принципы определения порогов, пригодные для работы со стационарными изображениями. На обработанном изображении наблюдаются в этом случае области, на которые разбито исходное изображение, отчетливо видны границы между областями. Это – существенный недостаток метода.

Более трудоемка, но и более эффективна процедура, использующая скользящее окно, при которой каждое новое положение рабочей области отличается от предыдущего только на один шаг по строке или по столбцу. Находимый на каждом шаге оптимальный порог относят к центральной точке текущей области. Таким образом, при этом методе порог изменяется в каждой точке кадра, причем эти изменения имеют характер, сопостави-

*a**б*

Рис. 7.2. Пример бинарной сегментации с автоматическим определением порога (см. также с. 203)



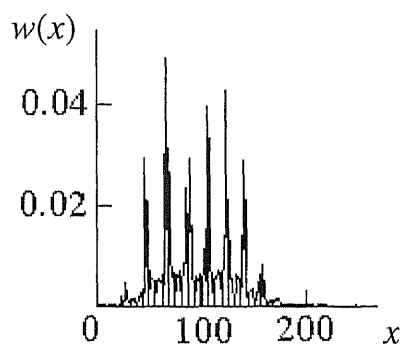
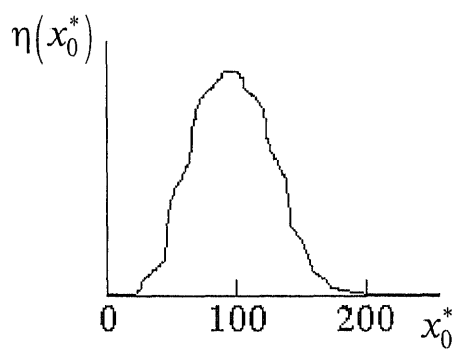
 $\theta$  $z$ 

Рис. 7.2. Окончание

мый с характером нестационарности самого изображения. Процедура обработки, конечно, существенно усложняется.

Компромиссной является процедура, при которой вместо скользящего окна с единичным шагом применяют “прыгающее” окно, перемещающееся на каждом этапе обработки на несколько шагов. В “пропущенных” точках кадра порог может определяться с помощью интерполяции (часто применяют простейшую линейную интерполяцию) по его найденным значениям в ближайших точках.

Оценивая результативность пороговой сегментации по рис. 7.2, б, следует отметить, что данный метод дает возможность получить определенное представление о характере однородных областей, образующих наблюдаемый кадр. Вместе с тем очевидно его принципиальное несовершенство, вызванное одноточечным характером принимаемых решений. Поэтому в последующих разделах обратимся к статистическим методам, позволяющим учитывать при сегментации геометрические свойства областей — размеры, конфигурацию и т.п. Отметим сразу же, что соответствующие геометрические характеристики задаются при этом своими вероятностными моделями и чаще всего в неявном виде.

## **7.2. СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ МАРКОВСКОЙ ФИЛЬТРАЦИИ**

Применение статистических методов фильтрации для сегментации изображений позволяет избавиться от основного недостатка простой пороговой обработки — ее одноточечного характера. Устоявшегося, безусловно предпочтительного метода сегментации на основе фильтрации в настоящее время не существует. Поэтому в данном и в следующем разделах будут рассмотрены два подхода, представляющихся наиболее перспективными для решения обсуждаемой задачи. Следует отметить, что теоретическая основа этих подходов значительно глубже, чем теория пороговой обработки. В пределах настоящего пособия будет дано лишь качественное изложение теории. При желании необходимые детали и доказательства могут быть найдены в публикациях.

Рассмотрим марковскую сегментацию, основанную на использовании неполных данных, которая по своей теоретической сути очень близка к марковской фильтрации изображений, рассмотренной в разделе 3.5 настоящего пособия. Статистическая сегментация, как и фильтрация, предполагает принятие решения

в каждой точке кадра на основе анализа исходного изображения в некоторой окрестности этой точки. В зависимости от характера этой окрестности сегментация может быть каузальной, полукаузальной и некаузальной. Здесь рассмотрим разновидность некаузальной сегментации, использующей лишь данные строки и столбца, на пересечении которых располагается текущая точка кадра, что иллюстрируется

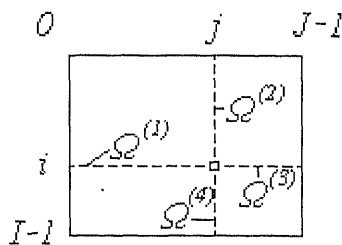


Рис. 7.3. Геометрия используемых данных при сегментации

(рис. 7.3). Напомним, что некаузальность в этом случае предполагает использование наблюдаемых данных всех четырех лучей  $\Omega^{(1)} \dots \Omega^{(2)} \dots$ , образующих в совокупности крестообразную структуру  $\Omega$ . Напомним также, что ограничение используемых данных лишь множеством, соответствующим рассматриваемой структуре, приводит к небольшой, как показывают детальные исследования, потере эффективности обработки. При обработке должны быть вычислены апостериорные вероятности всех классов, а затем принято решение об отнесении текущей точки кадра к тому классу, которому соответствует наибольшая вероятность.

Статистические методы всегда опираются на использование тех или иных математических моделей изображения. Ограничимся для простоты и здесь задачей бинарной сегментации. Рассмотрим случай, когда классы однородности отличаются друг от друга видом своих корреляционных функций, имея одинаковые средние яркости и дисперсии флюктуаций. На изображении (см., например, приводимый на рис. 7.4, а) участки разных классов различаются характером рисунка, размером деталей, величиной квазипериода. Такого рода различия относят к числу *текстурных*. Заметим, что понятие текстуры, достаточно широко применяемое в цифровой обработке изображений, до сих пор не имеет общепринятого определения. Многие определения ориентированы на присутствие повторяемости — строгой или приблизительной — элементов картины. Авторы [7.5] под текстурой понимают “описание пространственной упорядоченности элементов изображения в пределах некоторого участка изображения”.

Очевидна непригодность пороговой обработки для сегментации в рассматриваемом случае текстурных различий классов.

Будем считать, что в пределах отдельного участка однородности изображение описывается случайным полем с гауссовским распределением и биэкспоненциальной корреляционной функцией:

$$B_x(i, j) = D_x \cdot F^{|i|+|j|}, \quad (7.13)$$

где  $D_x$  — дисперсия,  $F$  — коэффициент корреляции соседних по строке или по столбцу элементов поля. Классы однородности в рассматриваемом случае корреляционных различий можно описать при помощи двух таких случайных полей  $x_0(i, j)$  и  $x_1(i, j)$  с разными значениями коэффициента одношаговой корреляции  $F_0$  и  $F_1$ . Среднее значение поля для обоих классов однородности здесь принято равным нулю, что не является определяющим. Наблюдаемое поле  $y(i, j)$  представим в виде

$$y(i, j) = x_\lambda(i, j), \quad (7.14)$$

где индекс  $\lambda$  в различных точках кадра случайным образом принимает значения 0 или 1. Тем самым в модели в каждой точке кадра выбор того или иного класса осуществляется при помощи управляющего случайного бинарного поля  $\lambda(i, j)$ , а задача сегментации формулируется как получение оценки  $\lambda^*(i, j)$  этого поля по наблюдаемому изображению (7.14).

Для того чтобы можно было применять описываемую ниже марковскую сегментацию, информационное поле  $\lambda(i, j)$  должно подчиняться определенному описанию. Доказано, что необходимым и достаточным условием применимости марковского аппарата является также биэкспоненциальный характер двумерной корреляционной функции. Одномерные сечения такого поля, получаемые в строках или в столбцах, представляют собой марковские последовательности. Одной из важнейших характеристик таких последовательностей является матрица одношаговых вероятностей перехода

$$\pi = \left\| \pi(\lambda_r | \lambda_{r-1}) \right\| = \begin{bmatrix} \pi(0|0) & \pi(1|0) \\ \pi(0|1) & \pi(1|1) \end{bmatrix}.$$

Ее элемент  $\pi(\lambda_r | \lambda_{r-1})$  — вероятность того, что на  $r$ -м шаге процесс примет значение  $\lambda_r$  при условии, что на предыдущем  $(r-1)$ -м шаге его значение равнялось  $\lambda_{r-1}$ . Матрица переходных

вероятностей, называемая стохастической, предусматривает все четыре возможные комбинации прошлого и текущего значений.

Обозначим  $\Lambda^{(k)}$ ,  $Y^{(k)}$  – векторы, образуемые значениями полей  $\lambda_{i,j} = \lambda(i, j)$  и  $y_{i,j} = y(i, j)$  соответственно в точках луча  $\Omega^{(k)}$  за исключением текущей точки  $(i, j)$ , а  $\Lambda$  и  $Y$  – аналогичные векторы, включающие все значения этих полей в точках “креста”  $\Omega$ . При принятых моделях полей совокупность векторов  $(\Lambda, Y)$  обладает марковским свойством условной независимости:

$$w(\Lambda, Y) = w(\lambda_{i,j}, y_{i,j}) \prod_{k=1}^4 w(\Lambda^{(k)}, Y^{(k)} | \lambda_{i,j}, y_{i,j}), \quad (7.15)$$

позволяющим компактно выразить совместное распределение вероятностей  $w(\Lambda, Y)$  через более простые условные распределения  $w(\Lambda^{(k)}, Y^{(k)} | \lambda_{i,j}, y_{i,j})$  на отдельных лучах. Важным достоинством представления (7.15) является вытекающая из него возможность определять апостериорное распределение вероятностей (АРВ)  $w(\lambda_{i,j} | Y)$  в виде

$$w(\lambda_{i,j} | Y) = c_{i,j} \frac{\prod_{k=1}^4 w(\lambda_{i,j} | Y^{(k)}, y_{i,j})}{w^3(\lambda_{i,j} | y_{i,j})}, \quad (7.16)$$

где  $c_{i,j}$  – нормировочная постоянная. По формуле (7.16) в каждой точке кадра должны быть вычислены две вероятности  $w(\lambda_{i,j} = 1 | Y)$  и  $w(\lambda_{i,j} = 0 | Y)$ , а затем вынесено решение  $\lambda_{i,j}^* = 1$  или  $\lambda_{i,j}^* = 0$  в зависимости от того, какая из этих вероятностей больше.

Удобство (7.16) состоит в том, что вычисление АРВ  $w(\lambda_{i,j} | Y)$  распадается на четыре однотипные независимые друг от друга вычислительные процедуры, выполняемые на каждом из лучей  $\Omega^{(k)}$ . Практически это означает, что для обработки всего кадра требуется проводить обработку каждой строки для получения каузальной  $w(\lambda_{i,j} | Y^{(1)}, y_{i,j})$  и антикаузальной  $w(\lambda_{i,j} | Y^{(3)}, y_{i,j})$  вероятностей. Технически же это реализуется двукратным сканированием строк – в прямом и в обратном направлениях. Аналогично двукратным сканированием столбцов вычисляют вероят-

ности  $w(\lambda_{i,j} | Y^{(2)}, y_{i,j})$  и  $w(\lambda_{i,j} | Y^{(4)}, y_{i,j})$ . Кроме того, в каждой точке кадра вычисляют одноточечное АРВ  $w(\lambda_{i,j} | y_{i,j})$ , входящее в (7.16).

Марковский характер случайных полей, описывающих изображение, приводит к еще одному важному достоинству вычислительных процедур, реализующих (7.16) — рекуррентности алгоритмов вычисления частных АРВ  $w(\lambda_{i,j} | Y^{(k)}, y_{i,j})$ . Это означает, что при движении вдоль  $k$ -го луча  $\Omega^{(k)}$  крестообразного множества  $\Omega$  АРВ в произвольной  $r$ -й точке этого луча определяется значением этой АРВ в предыдущей  $(r-1)$ -й точке. В целом, благодаря указанным качествам, алгоритм сегментации оказывается очень технологичным.

В качестве примера приведем без вывода рекуррентные уравнения для каузальной АРВ, т.е. описывающие процесс получения распределений на лучах  $\Omega^{(1)}$  и  $\Omega^{(2)}$ :

$$w(\lambda_r | y_0, \dots, y_r) = C_r \sum_{\lambda_{r-1}=0}^1 w(\lambda_{r-1} | y_0, \dots, y_{r-1}) \pi(\lambda_r | \lambda_{r-1}) \times \times w_x(y_r | y_{r-1}, \lambda_r, \lambda_{r-1}). \quad (7.17)$$

В выражении (7.17)  $C_r$  — нормирующий коэффициент;  $y_0, \dots, y_r$  — элементы изображения развернутой части строки  $\Omega^{(1)}$  или столбца  $\Omega^{(2)}$ , а  $w_x(y_r | y_{r-1}, \lambda_r, \lambda_{r-1})$  — распределение текущего элемента наблюдаемого изображения при его фиксированном значении на предыдущем шаге, а также при фиксированных значениях информационного сигнала на текущем и предыдущем шагах. Это условное распределение для принятой здесь математической модели определяется характеристиками поля  $x_\lambda(i, j)$  и описывается выражением:

$$w_x(y_r | y_{r-1}, \lambda_r, \lambda_{r-1}) = \frac{1}{\sqrt{2\pi D_x}} \exp \left( -\frac{(y_r - F_{\lambda_r} y_{r-1})^2}{2D_x(1 - F_{\lambda_r}^2)} \right),$$

в котором параметр  $F_{\lambda_r}$  равен  $F_0$  или  $F_1$  для  $\lambda_r$ , равного 0 или 1 соответственно. Как и (7.16), выражение (7.17) просчитывается для двух значений  $\lambda_r$ , равных 0 и 1, что в совокупности и составляет распределение вероятностей в случае бинарного информационного процесса.

Одноточечное АРВ  $w(\lambda_r | y_r)$ , участвующее в вычислении полного АРВ (7.16), находим по формуле Байеса:

$$w(\lambda_r | y_r) = c w_\lambda(\lambda_r) w_x(y_r | \lambda_r),$$

в которой  $c$  — нормировочный коэффициент для данного распределения,  $w_\lambda(\lambda_r)$  — априорное распределение вероятностей классов однородности. Одношаговая функция правдоподобия  $w_x(y_r | \lambda_r)$ , имеющая вид

$$w_x(y_r | \lambda_r) = \frac{1}{\sqrt{2\pi D_x}} \exp\left(-\frac{y_r^2}{2D_x}\right),$$

не зависит в данном случае от этого информационного поля, что обусловлено специфическим характером воздействия информационного поля на наблюдаемый сигнал. Следовательно, знаменатель в (7.16) совпадает с априорным распределением классов однородности изображения, а это исключает необходимость его вычислений в процессе сегментации.

Отметим, что вычисления на лучах  $\Omega^{(3)}$  и  $\Omega^{(4)}$  незначительно отличаются от вычислений, описанных для лучей  $\Omega^{(1)}$  и  $\Omega^{(2)}$ . Как было сказано, развертка на этих лучах выполняется в обратном направлении. Кроме того, необходимо учитывать, что переходные распределения вероятностей для обоих марковских полей  $y(i, j)$  и  $\lambda(i, j)$ , участвующих в математической модели, в обратном “времени” отличаются от соответствующих распределений прямого “времени” [7.2].

Читателя, желающего познакомиться с деталями выводов и доказательств, опущенных в данном изложении марковской сегментации, отсылаем к первоисточнику [7.2].

Приведем иллюстрации, демонстрирующие применение описанного метода. На рис. 7.4, а показано исходное текстурное изображение, на котором присутствуют области двух классов. Энергетические характеристики классов — средние яркости и дисперсии — одинаковы, коэффициент одношаговой корреляции  $F_0 = 0$ ,  $F_1 = 0,95$ . Сюжет “полуостров”, как видно из рис. 7.4, б, эффективно выделяется при помощи марковской процедуры сегментации.

Подчеркнем еще раз, что пороговая обработка в условиях текстурной сегментации абсолютно лишена смысла.

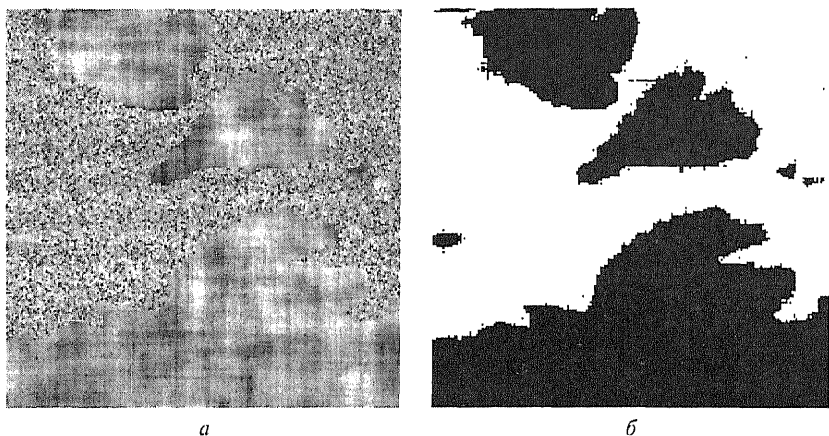


Рис. 7.4. Марковская сегментация текстурного изображения

### 7.3. БАЙЕСОВСКАЯ СЕГМЕНТАЦИЯ, ОСНОВАННАЯ НА РАСПРЕДЕЛЕНИИ ГИББСА

Описанная в предыдущем разделе марковская сегментация является одним из вариантов байесовского решения задачи. Здесь будет представлен еще один вариант, основу которого составляет так называемое распределение Гиббса. Его успешное применение к решению различных задач обработки изображений можно найти в работах [7.3, 7.4]. В этих работах, в частности, предложено и получило развитие решение задачи сегментации на основе распределения Гиббса.

Сущность байесовского метода сегментации, как говорилось в п.7.2, состоит в таком разбиении кадра на неперекрывающиеся области, которое в наибольшей степени соответствует наблюдаемому изображению. Мерой этого соответствия служит апостериорная вероятность того или иного варианта разбиения. Наиболее оправданным признается такое разбиение, апостериорная вероятность которого максимальна.

Особенностью рассматриваемого здесь байесовского метода является то, что он опирается на описание исходного изображения при помощи распределения Гиббса (РГ). Поэтому, прежде чем говорить о сегментации, необходимо познакомиться с основами использования РГ для вероятностного описания сигналов и изображений.



### 7.3.1. РАСПРЕДЕЛЕНИЕ ГИББСА И ЕГО ПРИМЕНЕНИЕ К ОПИСАНИЮ СЛУЧАЙНЫХ ДИСКРЕТНЫХ СИГНАЛОВ И ИЗОБРАЖЕНИЙ

Сразу отметим, что распределение Гиббса на самом деле представляет собой широкий класс распределений, объединяемых некоторыми общими фундаментальными чертами. По своей сути РГ может быть применено для описания *дискретных* сигналов и изображений. При решении статистических задач обработки сигналов или изображений одно из центральных мест принадлежит проблеме представления многомерного (совместного) распределения вероятностей всех отсчетов полезного сигнала и наблюдаемых данных. Особая трудность связана с заданием взаимозависимости элементов в этом многомерном распределении. Распределение Гиббса при решении этой проблемы опирается на геометрическое понятие *клик* и понятие *потенциалов взаимодействия* значений сигнала на этих кликах. Поясним их на примере случайного поля.

Элемент двумерного сигнала (изображения)  $x_{i,j}$  в точке кадра с координатами  $i, j$  связан (взаимодействует) с другими элементами, что обуславливает их вероятностную взаимозависимость. Взаимодействия могут включать два и более элемента изображения, в первом случае взаимодействия называются парными. Вокруг каждой точки изображения формируется несколько групп точек, в каждую из которых входит и данная точка. Каждая из этих групп есть клика. На рис. 7.5 показан пример множества кликов, включающих точку  $i, j$  изображения. Множество содержит восемь кликов, объединяемых овалами. Каждая состоит из двух точек: кроме самой точки  $i, j$  в них входит по одной из точек ближайшего окружения.

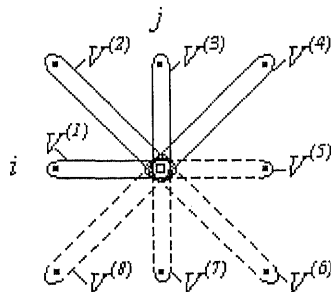


Рис. 7.5. К понятию клик  
и потенциалов взаимодействия

Клики, показанные на рис. 7.5, относятся к одному из четырех типов — горизонтальному, вертикальному или одному из наклонных. Точка  $i, j$  входит в две клики каждого типа. Для более компактной записи последующих соотношений половину из показанных на рис. 7.5 клик (выделены непрерывными линиями) закрепляют за точкой  $i, j$ . Аналогичное правило закрепления

клик применяется и для всех других точек. В результате все парные клики, существующие на двумерной решетке, оказываются однозначно распределенными между точками изображения.

Таким образом, вся решетка, на которой определено изображение, оказывается связанной множеством парных кликов. Геометрическая структура кликов может быть самой разнообразной. В них могут входить не только ближайшие по решетке, но и удаленные соседи, клики могут различаться числом входящих в них точек, одновременно в модели могут присутствовать более простые и более сложные клики и т.д. Многообразие геометрических вариантов представляет собой один из источников разнообразия математических моделей изображений, основанных на РГ.

На каждой клике задается *потенциал взаимодействия* (ПВ) — функция значений изображения в точках, составляющих клику. На рис. 7.5  $V^{(i)}$ ,  $i = \overline{1, 8}$  — множество потенциалов. Для однородного изображения, т.е. такого, у которого порождающий его механизм одинаков на всем кадре, очевидно справедливы равенства функций:

$$V^{(1)} = V^{(5)}, V^{(2)} = V^{(6)}, V^{(3)} = V^{(7)}, V^{(4)} = V^{(8)}. \quad (7.18)$$

Пусть  $x_{i,j}$  — бинарное поле, т.е. его значения в разных точках равны 0 или 1. В этом случае каждый из потенциалов представляет собой функцию, которая может быть задана при помощи табл. 7.1. Здесь в качестве примера приведена функция ПВ

Таблица 7.1  
Потенциал  $V^{(1)}$  парного взаимодействия  
бинарного поля

| $x_{i,j-1}$ | $x_{i,j}$       |                 |
|-------------|-----------------|-----------------|
|             | 0               | 1               |
| 0           | $V_{0,0}^{(1)}$ | $V_{1,0}^{(1)}$ |
| 1           | $V_{0,1}^{(1)}$ | $V_{1,1}^{(1)}$ |

$V^{(1)}$  (рис. 7.5), аргументами которой являются отсчеты изображения  $x_{i,j}$  и  $x_{i,j-1}$ . В этом случае ПВ описывается всего четырьмя числами, составляющими данную таблицу. Поскольку модель однородного поля в соответствии с (7.18) задается четырьмя потенциалами  $V^{(1)} \dots V^{(4)}$ , то ее полное задание требует нахождения 16 чисел. Если изображение обладает некоторой симметрией, то количество независимых потенциалов, входящих в соответствующие таблицы, будет меньшим. Так, если свойства изображения вдоль строк и вдоль столбцов одинаковы, то  $V^{(1)} = V^{(3)}$ , а при совпадении свойств по диаго-

дает некоторой симметрией, то количество независимых потенциалов, входящих в соответствующие таблицы, будет меньшим. Так, если свойства изображения вдоль строк и вдоль столбцов одинаковы, то  $V^{(1)} = V^{(3)}$ , а при совпадении свойств по диаго-

нальным направлениям  $V^{(2)} = V^{(4)}$ . При этих условиях математическая модель полностью определяется 8 параметрами.

С учетом введенных обозначений распределение Гиббса для данного примера имеет вид

$$P(X) = c_x \exp \left\{ - \sum_{i,j \in \Omega_x} \left[ V_{x(i,j),x(i,j-1)}^{(1)} + V_{x(i,j),x(i-1,j-1)}^{(2)} + \right. \right. \\ \left. \left. + V_{x(i,j),x(i-1,j)}^{(3)} + V_{x(i,j),x(i-1,j+1)}^{(4)} \right] \right\}, \quad (7.19)$$

где  $X$  — совокупность всех отсчетов изображения,  $\Omega_x$  — множество узлов решетки, на котором задано изображение, а  $c_x$  — константа, которую находят из условия нормировки распределения вероятностей:

$$\sum P(X) = 1, \quad (7.20)$$

в котором суммирование выполняется по множеству реализаций случайного поля. Обычно РГ записывают в более компактной форме

$$P(X) = c_x \exp \{-U(X)\}, \quad (7.21)$$

где вводят понятие потенциальной функции

$$U(X) = \sum_{i,j \in \Omega_x} \sum_{k=1}^4 V_x^{(k)}, \quad (7.22)$$

а отдельные (частные) потенциалы  $V_x^{(k)}$  представляют в общей записи как функции всего изображения  $X$  (хотя в действительности каждый из них, как мы видели выше, зависит лишь от определенной пары отсчетов).

Во втором примере, который мы рассмотрим, случайное поле  $y_{i,j}$  принимает значения из непрерывного множества  $-\infty < y_{i,j} < \infty$ . Оно подчиняется распределению Гиббса, для которого клики имеют такой же вид, что и в примере бинарного поля (рис. 7.5), а потенциалы взаимодействия (на примере потенциала  $V_y^{(1)}$ ) — квадратичные функции вида

$$V_y^{(1)} = (y_{i,j} - \rho^{(k)} y_{i,j-1} - m)^2 / 2D, \quad (7.23)$$

с параметрами  $m$ ,  $D$  и  $\rho^{(k)}$ . Обозначим  $y_{i,j}^{(k)}$  — элемент поля, соответствующий  $k$ -й клике для центрального элемента с координатами  $i, j$ . Например,  $y_{i,j}^{(1)} = y_{i,j-1}$ ,  $y_{i,j}^{(2)} = y_{i-1,j-1}$  и т.д. Тогда в развернутом виде РГ будет иметь вид:

$$P(Y) = c_y \exp \left\{ - \sum_{i,j \in \Omega_y} \sum_{k=1}^4 (y_{i,j} - \rho^{(k)} y_{i,j}^{(k)} - m)^2 / 2D \right\}. \quad (7.24)$$

В компактной же форме остается справедливой запись вида (7.21), но с потенциалами (7.23). Константа  $c_y$  также определяется выражением типа (7.20), однако вместо многомерного суммирования в нем использовано интегрирование. Аргумент экспоненты в (7.24) — квадратичная форма, следовательно, данное РГ гауссовское.

Используя рассмотренные примеры однородных полей  $x_{i,j}$  и  $y_{i,j}$ , можно образовать математическую модель неоднородного изображения. Полагаем, что наблюдаемым является поле  $y_{i,j}$ , которое описывается РГ (7.24), но его параметры  $m$ ,  $D$  и  $\rho^{(k)}$  могут принимать одно из двух значений  $m_0$ ,  $D_0$  и  $\rho_0^{(k)}$  или  $m_1$ ,  $D_1$  и  $\rho_0^{(1)}$ . Выбор того или иного варианта в точке  $i, j$  определяется значением ненаблюдаемого бинарного поля  $x_{i,j}$ , которое оно принимает в данной точке. Таким образом, поле  $x_{i,j}$  является *управляющим, скрытым* в наблюдаемом изображении  $y_{i,j}$ , поскольку оно скачкообразно переводит это наблюдаемое изображение из одного состояния в другое. При этом сегментация заключается в получении оценки  $x_{i,j}^*$  ненаблюдаемого бинарного поля путем анализа наблюдаемого  $y_{i,j}$ .

Использование статистического подхода к сегментации требует, как будет видно из дальнейшего, использования совместного распределения  $P(X, Y)$  двух полей, для которого справедливо соотношение

$$P(X, Y) = P(X)P(Y|X),$$

Здесь  $P(X)$  — распределение множества всех значений управляющего поля  $x_{i,j}$ , а  $P(Y|X)$  — условное по отношению к его

значению  $X = \{x_{i,j}\}$  распределение наблюдаемого изображения  $Y = \{y_{i,j}\}$ . Первое описывается соотношением (7.19) или (7.21) – (7.22), а для второго можно использовать (7.24), введя в него зависимость параметров  $m$ ,  $D$  и  $\rho^{(k)}$  от  $X$ . В результате имеем:

$$P(X, Y) = c \exp \left\{ - \sum_{i,j} \sum_{\Omega} \sum_{k=1}^4 \left[ V_x^{(k)} + \frac{(y_{i,j} - \rho^{(k)}(x_{i,j}) y_{i,j}^{(k)} - m(x_{i,j}))^2}{2D(x_{i,j})} \right] \right\}, \quad (7.25)$$

где  $\Omega$  – множество узлов решетки, общее для обоих случайных полей, а  $c$  – нормировочная константа, обеспечивающая равенство единице суммарной вероятности для дискретного поля  $x_{i,j}$  и непрерывного  $y_{i,j}$ .

### 7.3.2. БАЙЕСОВСКАЯ СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЯ НА ОСНОВЕ СТОХАСТИЧЕСКОЙ РЕЛАКСАЦИИ

Оптимальная сегментация, основанная на байесовском принципе, должна вырабатывать такой результат  $x_{i,j}^*$ , которому соответствует максимум апостериорного распределения вероятностей (АРВ):

$$X^* = \arg \max_X P(X|Y), \quad (7.26)$$

где  $X^* = \{x_{i,j}^*\}$  – результат оптимальной сегментации. Располагая совместным распределением (7.25), можно найти и АРВ, поскольку

$$P(X|Y) = \frac{P(X, Y)}{P(Y)}. \quad (7.27)$$

Однако вычисления (7.27) можно избежать. В самом деле, АРВ  $P(X|Y)$  и совместное распределение  $P(X, Y)$  различаются лишь множителем  $P(Y)$ , который не зависит от  $X$ . Поэтому характер зависимости от  $X$  этих двух распределений одинаков, следовательно, вектор  $X^*$ , максимизирующий АРВ, доставляет максимум совместному распределению и, наоборот, если  $X^*$  максимизирует  $P(X, Y)$ , то при этом максимизируется и  $P(X|Y)$ . Эти

соображения позволяют вместо требования (7.26) для отыскания оптимального результата  $X^*$  использовать требование

$$X^* = \arg \max_X P(X, Y). \quad (7.28)$$

Оценим практическую возможность реализации процедуры (7.28). Пусть кадр  $\Omega$  содержит  $I$  строк и  $J$  столбцов, так что общее число узлов равно  $IJ$ , а количество различных бинарных полей  $x_{i,j}$  определяется числом  $N = 2^{I \cdot J}$ . В этом случае распределение  $P(X, Y)$ , рассматриваемое как функция  $X$   $f(X) = P(X, Y)|_{Y=Y^{(H)}}$  при фиксированном  $Y = Y^{(H)}$ , можно представить в виде таблицы, содержащей  $N$  значений. Здесь  $Y^{(H)}$  — наблюдаемое изображение. Поиск решения (7.28) предполагает вычисление значений функции  $f(X)$  и определение точки ее максимума  $X^*$ . Если, например,  $I = J = 256$ , то размер таблицы  $N = 2^{256 \times 256} = 2^{65536} \cong 10^{20000}$ . Понятна бесперспективность такого способа решения даже при бинарной сегментации.

Как альтернатива данному традиционному методу получения байесовских оценок развивается теория, получившая название *стохастической релаксации*. Вместо полного перебора вариантов, описанного выше, в этом случае осуществляют целенаправленное вероятностное моделирование случайного поля с использованием итерационной процедуры. На  $r$ -м шаге итерационного процесса формируется поле  $x_{i,j}^{(r)}$ , а основой его генерирования являются результат предыдущего шага  $x_{i,j}^{(r-1)}$  и распределение  $P(X, Y)|_{Y=Y^{(H)}}$ . При рационально организованном итерационном процессе имеет место сходимость  $x_{i,j}^{(r)}$  к  $x_{i,j}^*$ , причем, хотя количество итераций оказывается достаточно большим, общее число вычислений получается вполне приемлемым. Существует несколько методов стохастической релаксации, реализующих эту идею. Рассмотрим один из них, так называемый *метод Метрополиса* [7.3, 7.4], более подробно.

На первом шаге итерационного процесса ( $r=1$ ) случайное поле  $x_{i,j}^{(1)}$  формируют при помощи датчика случайных чисел с равномерным распределением вероятностей на интервале  $[0, 1]$ . Для этого данный интервал разбивают на два подынтервала, а в качестве граничной точки используют значение вероятности  $P_0$

того, что  $x_{i,j}=0$ . При неизвестном значении  $P_0$  его принимают равным 0,5. При помощи датчика формируют число  $\xi_{i,j}$ , которое затем путем сравнения с порогом  $P_0$  преобразуют в  $x_{i,j}^{(1)}$  в соответствии с правилом:

$$x_{i,j}^{(1)} = \begin{cases} 0, & \text{при } \xi_{i,j} \leq P_0 \\ 1, & \text{при } \xi_{i,j} > P_0. \end{cases}$$

Отметим, что получаемое в результате первого шага случайное поле состоит из независимых значений в силу независимости чисел, генерируемых датчиком.

Второй и последующие шаги итерационного процесса принято называть *макрошагами*. Макрошаг состоит из  $IJ$  *микрошагов* — по числу точек изображения. При выполнении одного микрошага одна из точек кадра находится в состоянии *текущей точки*, в результате же одного макрошага через это состояние поочередно проходят все  $IJ$  точек. При этом авторы метода рекомендуют использовать случайный порядок обхода кадра для улучшения сходимости релаксационного процесса.

На каждом микрошаге решается вопрос о возможном изменении только того отсчета  $x_{i,j}^{(r)}$ , который является текущим. Для этого сначала, как и на первом шаге, при помощи датчика случайных чисел вырабатывается число  $\tilde{x}_{i,j}^{(r)}$  (которое может совпадать с прежним значением поля  $x_{i,j}^{(r-1)}$  в этой точке или отличаться от него). В результате формируется промежуточный вектор  $\tilde{X}^{(r)}$ , который, возможно, отличается от  $X^{(r-1)}$  только в одной текущей точке с координатами  $i, j$ . Затем по формуле (7.25) подсчитываются вероятности  $\tilde{P} = P(\tilde{X}^{(r)}, Y^{(H)})$  и  $P = P(X^{(r-1)}, Y^{(H)})$ , сравнение которых позволяет определить, какой из векторов  $\tilde{X}^{(r)}$  и  $X^{(r-1)}$  является более вероятным. Если  $\tilde{P} > P$  (это может быть лишь при различных значениях  $\tilde{x}_{i,j}^{(r)}$  и  $x_{i,j}^{(r-1)}$ ), то за результат этого шага принимается вектор  $X^{(r)} = \tilde{X}^{(r)}$ . Следовательно, при таком развитии процесса на данном микрошаге происходит изменение одного (текущего) элемента поля.

Если же  $\tilde{P} \leq P$ , то вычисляется отношение этих вероятностей  $\gamma = (\tilde{P}/P) \leq 1$ , которое используется в качестве граничной величины при еще одном обращении к датчику равновероятных чисел.

Пусть  $\xi_{i,j}^{(r)}$  — результат работы датчика. Тогда результат данного микрошага формируется по правилу:

$$x_{i,j}^{(r)} = \begin{cases} \tilde{x}_{i,j}^{(r)}, & \text{при } \xi_{i,j}^{(r)} < \gamma \\ x_{i,j}^{(r-1)}, & \text{при } \xi_{i,j}^{(r)} \geq \gamma. \end{cases}$$

Таким образом, согласно этому правилу с определенными шансами, зависящими от числа  $\gamma$ , может произойти как изменение значения поля в текущей точке, так и его сохранение.

Отметим, что в процессе вычисления вероятностей  $\tilde{P}$  и  $P$  по формуле (7.25) не требуется предварительно находить нормировочную константу  $c$ , что было бы чрезвычайно трудоемкой вычислительной процедурой, переводящей метод стохастической релаксации за грань реализуемости. Данная возможность связана с использованием этих вероятностей только для сравнения друг с другом. Поэтому можно, например, принять  $c = 1$ .

Отметим также, что гиббсовский характер распределения (7.25) позволяет существенно облегчить вычисления при текущем расчете вероятностей  $\tilde{P}$  и  $P$ . В многомерной сумме, входящей в это выражение, на каждом микрошаге изменяются лишь те слагаемые, которые через клики связаны с текущей точкой  $i, j$  кадра. Поэтому процедура должна быть организована так, чтобы выполнялись лишь эти минимально необходимые вычисления при максимальном использовании рассчитанных ранее результатов.

Важной задачей при создании любого итерационного вычислительного процесса является обеспечение его сходимости. Для этой цели используют несколько модифицированный вид распределения (7.25), в которое вводят параметр  $\lambda^{(r)}$ , управляющий сходимостью. Модифицированное распределение имеет вид:

$$P(X, Y) = c \exp \left\{ -\lambda^{(r)} U_{\Sigma}(X, Y) \right\},$$

где

$$U_{\Sigma}(X, Y) = \sum_{i,j \in \Omega} \sum_{k=1}^4 \left[ V_x^{(k)} + \frac{(y_{i,j} - \rho^{(k)}(x_{i,j}) y_{i,j}^{(k)} - m(x_{i,j}))^2}{2D(x_{i,j})} \right] -$$

энергетическая функция. Нетрудно понять, что при увеличении параметра  $\lambda^{(r)}$  происходит обострение распределения  $P(X, Y)$  в окрестности его пиков. Поэтому в ходе итерационного процесса значение этого параметра изменяют, увеличивая при переходе к каждому новому макрошагу. Выбор скорости изменения являет-



ся очень сложной задачей. При быстром увеличении будет повышаться скорость сходимости итерационного процесса, однако при этом возрастает опасность “застревания” в окрестности какого-либо локального максимума функции  $P(X, Y)$ . Экспериментально установлено, что эти противоречия преодолеваются, если увеличение параметра подчиняется логарифмической зависимости вида  $\lambda^{(r)} \approx \ln(1 + r)$ .

На рис. 7.6 показаны результаты эксперимента по бинарной сегментации реального изображения участка земной поверхности, выполненного на базе описанного метода.<sup>1</sup> На рис. 7.6, а показано исходное изображение, а на рис. 7.6, б и 7.6, в — результаты его бинарной сегментации на основе порогового (порог равен 180) и гиббсовского методов соответственно. Очевидно, что гиббсовская сегментация субъективно воспринимается как более совершенная.

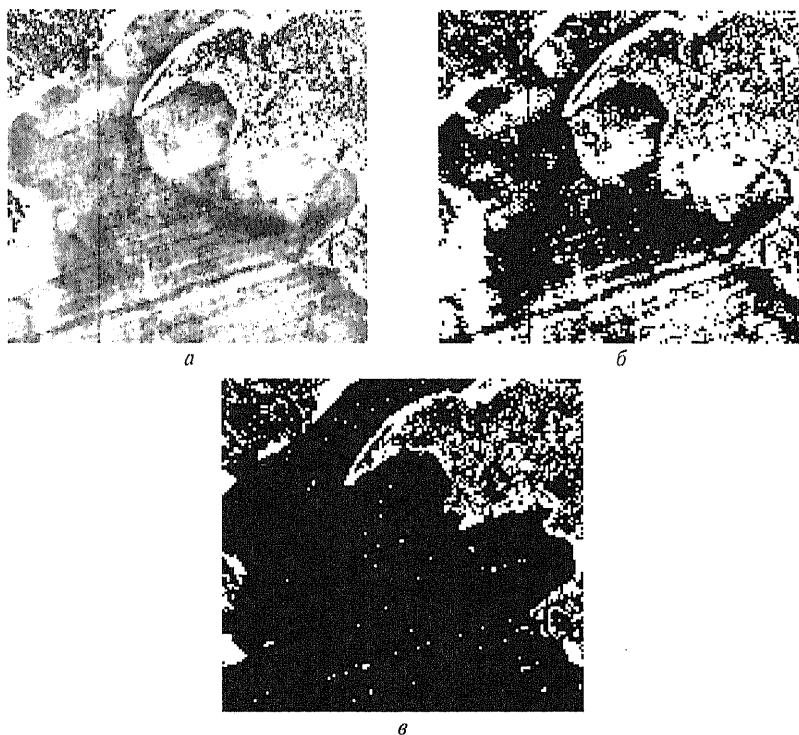


Рис. 7.6. Гиббсовская сегментация изображения

<sup>1</sup> Эксперимент выполнен магистрантом А.В. Юриным под руководством доц. В.Н. Васюкова

Критерий прекращения итерационного процесса основывается на анализе степени обновления результата данного макрошага  $X^{(r)}$  по сравнению с предыдущим макрошагом  $X^{(r-1)}$ . На первых макрошагах количество отсчетов поля, изменяющих свое значение, велико. Постепенно это число уменьшается, стремясь в пределе к нулю, что и означает сходимость итерационного процесса. Поэтому можно использовать правило, при котором процедура прекращается после очередного макрошага, если процент обновившихся точек не превысил заданной величины  $\epsilon$ . Понятно, что продолжительность процедуры является случайной величиной. В проведенном эксперименте величина  $\epsilon$  была выбрана равной 0,01, а продолжительность процедуры составила 235 макрошагов.

В заключение укажем на существенную трудность применения гиббсовских методов обработки сигналов и изображений, связанную с измерением параметров распределения Гиббса для достижения адекватности математического описания. В настоящее время подбор параметров осуществляют путем вероятностного моделирования случайного поля и сравнения получаемой картины с реальным изображением. Этот метод не может считаться удовлетворительным в силу своей субъективности. Разработка эффективных методов идентификации гиббсовских моделей является задачей будущих исследований.

## ВОПРОСЫ К ГЛАВЕ 7

7.1. Какие свойства реальных изображений препятствуют определению порогов сегментации по “провалам” в гистограммах?

7.2. Поясните смысл дискриминантного критерия, применяемого для автоматического определения порогов при яркостной сегментации изображений.

7.3. Докажите справедливость соотношений (7.10), (7.11) и (7.12), составляющих основу алгоритма автоматического определения порога сегментации.

7.4. Какая математическая модель изображения используется при марковской сегментации?

7.5. В чем состоят основные черты марковской сегментации изображений при использовании неполных данных?

7.6. Поясните структуру процесса вычисления частных АРВ на отдельных лучах  $\Omega^{(k)}$  при марковской сегментации.

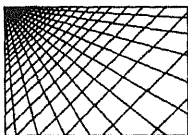
7.7. Поясните, почему при марковской сегментации изображения по текстурному признаку одноточечное АРВ совпадает с априорным распределением классов однородности.

7.8. Поясните, что понимается под кликами и какова их роль в теории распределения Гиббса.

7.9. Какой вид имеет распределение Гиббса, как при его помощи можно описать однородное изображение?

7.10. Как при помощи распределения Гиббса можно построить математическую модель неоднородного изображения для решения задачи сегментации?

7.11. В чем состоит назначение и сущность стохастической релаксации?



## ГЛАВА 8

### ВЫДЕЛЕНИЕ КОНТУРНЫХ ЛИНИЙ

Одной из основных и важнейших целей цифровой обработки изображений является распознавание присутствующих на них объектов. Возможность различать объекты заложена в высокой информативности изображения. В значительной мере именно это и привлекает разработчиков различных информационных систем к использованию изображений как способу представления результатов наблюдения.

Вместе с тем предъявляемые к обработке изображения содержат много избыточных малоинформативных сведений, которые занимают, однако, большие объемы памяти, и требующих выполнения большого количества вычислений при попытке использовать их для распознавания. В теории распознавания образов существуют методы, позволяющие выяснять степень информативности тех или иных признаков. Однако в настоящее время эти методы не нашли широкого применения в распознавании изображений.

Значительно шире применяются методы сокращения избыточности, опирающиеся на специфические особенности зрительного восприятия изображений. Считается, что субъективное восприятие наблюдаемой сцены происходит через ее представление в виде отдельных однородных областей (т.е. сегментацию) и выделение контурных линий. Контурные, или граничные, линии разделяют на изображении участки с различными свойствами, поэтому выделение контуров иногда рассматривается как предварительная обработка, направленная на последующее выполнение сегментации. Вместе с тем, препарат, образующийся в результате выделения контурных линий, в некоторых случаях может и самостоятельно эффективно использоваться для распознавания, поскольку содержащаяся в нем информация, по край-

ней мере, с точки зрения зрительного восприятия, вполне достаточно для решения многих задач такого типа.

Названные причины послужили стимулом для разработки методов выделения контуров, которым в современной теории цифровой обработки изображений уделяется большое внимание. Данная глава знакомит читателей с основными подходами к решению этой задачи.

Традиционное решение задачи выделения контуров включает в себя четыре этапа: подчеркивание (усиление) яркостных перепадов, выделение граничных точек, утоньшение (скелетизация) и устранение разрывов. Однако обычно лишь двум первым уделяется значительное внимание. Решение этих проблем будет предметом и нашего рассмотрения.

Развиваются и новые подходы, основанные на применении статистически оптимальных методов, способных работать в условиях, когда, например, средняя яркость изображения заранее неизвестна либо может изменяться в пределах рабочего поля. Один из таких методов также рассмотрен ниже.

### 8.1. ГРАДИЕНТНЫЕ МЕТОДЫ ПОДЧЕРКИВАНИЯ КОНТУРОВ

Смысл операции подчеркивания контуров состоит в том, чтобы усилить резкие перепады яркости, сформировав в соответствующих точках кадра импульсные отклики на фоне сравнительно слабых флуктуаций в других областях картины, не содержащих крутых яркостных перепадов. По своему характеру все операторы подчеркивания контуров являются различными модификациями дифференцирования двумерного поля по различным направлениям области определения. Как известно, при дифференцировании, в самом деле, в окрестности перепада функции яркости образуется пик, способствующий регистрации этой области. Однако также хорошо известно, что при дифференцировании сигналов с помехами происходит существенное усиление последних, что снижает отношение сигнал/шум и негативно отражается на получаемых результатах. Поэтому есть значительное число операторов, представляющих различные варианты дифференцирования с накоплением, применяемым для ослабления отрицательных последствий дифференцирования.

Обычно оператор подчеркивания контуров представляют в форме масочного линейного фильтра. В процессе обработки маска, которая является просто матрицей коэффициентов, скользит по полю изображения, занимая поочередно все возможные положения. В каждом положении маска играет роль ок-

на, при помощи которого отбираются отсчеты обрабатываемого изображения и выполняется их поэлементное умножение на соответствующий элемент маски с последующим суммированием всех произведений. Полученное число рассматривается как отсчет выходного изображения в точке, соответствующей центру симметрии окна.

Приведем часто используемые варианты масочных операторов подчеркивания. Простейшие операторы дифференцирования вдоль строки, столбца и двух диагональных направлений соответственно описываются матрицами:

$$\| -1, 1 \|, \begin{vmatrix} -1 \\ 1 \end{vmatrix}, \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}, \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} \quad (8.1)$$

Видно, что вычисление производной заменяется взятием разности элементов. Так, например, первый из приведенных операторов вычисляет разность вида  $y(i, j+1) - y(i, j)$ , что является аппроксимацией производной изображения  $y(i, j)$  по строке. Аналогично работают и остальные операторы, представленные в (8.1). Отметим, что два последних из них, работающих вдоль диагональных направлений, называются оператором Робертса.

Всем операторам в (8.1) свойственно два существенных недостатка. Первый — отсутствие накопления, что делает их, как уже говорилось, весьма чувствительными к шуму, содержащемуся в изображении  $y(i, j)$ . Второй состоит в том, что вычисляемые разности центрированы относительно средних точек между узлами. Так, например, первый оператор вычисляет производную в точке  $(i, j+1/2)$ . Таким образом, получаемое в результате изображение оказывается смещенным на половину шага относительно исходного, что может создать нежелательные проблемы при последующей обработке.

В приводимых ниже операторах чувствительность к шуму ослабляется благодаря увеличению размеров матриц, а смещение устраняется за счет использования масок с нечетным размером сторон. При этом обычно вычисляемый результат соответствует центральной точке окна. Наиболее часто применяют операторы Превитт I и II, матрицы которых имеют вид соответственно:

$$A_{P1} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{vmatrix} \text{ и } A_{P2} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{vmatrix} \quad (8.2)$$

и Собеля: 
$$\mathbf{A}_S = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \quad (8.3)$$

Приведенные здесь операторы вычисляют сглаженные производные вдоль столбцов изображения. В самом деле, в точке с координатами  $i, j$ , применяя, например, оператор Превитт-I, имеем

$$\dot{y}_v(i, j) = [y(i-1, j-1) - y(i+1, j-1)] + [y(i-1, j) - y(i+1, j)] + [y(i-1, j+1) - y(i+1, j+1)], \quad (8.4)$$

где через  $\dot{y}_v(i, j)$  обозначена производная вдоль вертикального направления. Согласно (8.4) получаемый результат представляет собой сумму трех разностей, взятых вдоль соседних столбцов изображения.

Поворотом весовых коэффициентов этих операторов на  $90^\circ$  можно получить операторы, вычисляющие сглаженную производную  $\dot{y}_r(i, j)$  изображения в горизонтальном направлении. Если же выполнить поворот весовых коэффициентов в матрицах (8.2), (8.3) на  $+45^\circ$  или  $-45^\circ$ , то получим операторы, вычисляющие сглаженные производные  $\dot{y}_\lambda(i, j)$  или  $\dot{y}_\gamma(i, j)$  вдоль двух диагональных направлений.

Известны различные способы дальнейшего использования производных, вычисленных при помощи масочных операторов, однако все они могут быть отнесены к градиентным методам. Градиент  $\nabla y(i, j)$  двумерной функции  $y(i, j)$  представляет собой, как известно, векторную двумерную функцию, определяющую направление и скорость максимального изменения функции  $y(i, j)$ . Чаще пользуются второй из этих характеристик, определяемой модулем градиента:

$$h(i, j) = |\nabla y(i, j)| = \sqrt{\left(\frac{\partial y(i, j)}{\partial i}\right)^2 + \left(\frac{\partial y(i, j)}{\partial j}\right)^2}. \quad (8.5)$$

При практическом нахождении производных, входящих в (8.5), применяют те или иные варианты их вычисления с помощью масочных операторов. Вычисление угла, указывающего направление максимального изменения функции яркости, выполняется на основе выражения:

$$\Theta(i, j) = \arctg \left[ \frac{\partial y(i, j)}{\partial i} / \frac{\partial y(i, j)}{\partial j} \right]. \quad (8.6)$$

При практической реализации (8.6) также прибегают к использованию масочных операторов.

Более грубыми, но все чаще применяемыми в практических задачах являются модификации градиентного метода, основанные на выборе максимальной из производных по различным направлениям и ее использовании в качестве градиента. Пусть

$$\dot{y}_{e_m}(i, j) = \frac{\partial y(i, j)}{\partial e_m} -$$

производная функции яркости в направлении  $e_m$ . Предположим, что используется  $M$  различных направлений. Тогда в качестве оценки градиента можно использовать ту из производных, модуль которой имеет максимальное значение:

$$|\nabla y(i, j)| \approx \max_{m=1, M} |\dot{y}_{e_m}(i, j)|.$$

Естественно, оценка угла  $\Theta(i, j)$  в этом случае определяется ориентацией соответствующей маски.

Кроме приведенных масочных операторов для вычисления сглаженных производных применяется и ряд других. Представляет интерес определить общие требования к этим операторам. Для этого представим матрицу масочного оператора дифференцирования со сглаживанием в общем виде:

$$A = \|a_{i,j}\| = \begin{vmatrix} A & B & C \\ D & E & F \\ G & H & V \end{vmatrix}. \quad (8.7)$$

Следуя [8.1], приведем некоторые из этих требований.

1. Матрицы масочных операторов, соответствующих горизонтальному и вертикальному направлениям, должны быть эквивалентны друг другу, отличаясь лишь поворотом своих элементов на  $90^\circ$ . Это свойство, в частности, дает возможность обсуждать все остальные свойства только для одного из направлений перепадов яркости, например, горизонтального, для подчеркивания которого требуется выполнять дифференцирование вдоль столбцов изображения.



2. Элементы матрицы должны быть симметричны относительно вертикальной оси матрицы

$$A = C, \quad D = F, \quad G = V, \quad (8.8)$$

поскольку соответствующие им элементы изображения равноценны.

3. Элементы, расположенные симметрично относительно горизонтальной оси, должны иметь противоположные знаки:

$$A = -G, \quad C = -V, \quad B = -H. \quad (8.9)$$

Этим условием гарантируется одинаковость абсолютной величины реакции масочного оператора на данный перепад и на перепад, получаемый из данного его зеркальным отражением относительно горизонтальной оси симметрии маски.

4. Производная постоянного сигнала должна равняться нулю. Из этого следует условие:

$$A + B + C + D + E + F + G + H + V = 0, \quad (8.10)$$

откуда с учетом (8.9) имеем:

$$D + E + F = 0. \quad (8.11)$$

Объединяя свойства (8.8)...(8.11), приводим матрицу (8.7) к виду:

$$A = \begin{vmatrix} A & B & A \\ D & -2D & D \\ -A & -B & -A \end{vmatrix}. \quad (8.12)$$

Запишем выражение для реакции матричного оператора (8.12) на входное изображение, представив для наглядности его фрагмент в произвольно выбранном окне также в виде матрицы:

$$\mathbf{y} = \|y_{i,j}\| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & v \end{vmatrix}.$$

Имеем:

$$\frac{\partial y(i,j)}{\partial i} = A(a-g) + A(c-v) + B(b-h) + \\ + D(d-e) + D(f-e) \quad (8.13)$$

В данном выражении три первых слагаемых имеют смысл конечных разностей, взятых в вертикальном направлении, что со-

ответствует вычислению производной вдоль столбцов изображения. Двум же последним слагаемым придать такой смысл не удастся, поэтому представляется естественным принять  $D$  равным нулю. В результате матрица (8.12) приобретает вид:

$$A = \begin{bmatrix} A & B & A \\ 0 & 0 & 0 \\ -A & -B & -A \end{bmatrix}. \quad (8.14)$$

Сопоставляя приведенные выше матрицы операторов Превитта – I и II и Собеля с полученным каноническим представлением (8.14), отмечаем, что матрица  $A_{P2}$  оператора Превитт-II не удовлетворяет последнему требованию. Следует отметить, что известны и другие применяемые на практике матричные операторы, содержащие те или иные отклонения от канонической формы.

Многообразие масочных операторов ставит вопрос о сравнении их характеристик и о выборе лучшего в том или ином смысле оператора. Для большинства известных в настоящее время операторов сопоставительный анализ выполняли различные авторы, а результаты этого анализа обсуждаются в [8.1]. Приведем основные выводы, содержащиеся в этой обзорной работе.

Для сравнения масочных операторов используется несколько показателей качества. Одним из основных является чувствительность оператора к ориентации яркостного перепада в кадре, оцениваемая величиной

$$\alpha(\Theta) = h(i, j|\Theta) / h(i, j|\Theta = 0). \quad (8.15)$$

Здесь  $h(i, j|\Theta)$  — значение градиента, вычисленное по формуле (8.5) при условии, что яркостный перепад проходит точно через центр симметрии маски и характеризуется наклоном  $\Theta$  к горизонтальному направлению. Таким образом, в соответствии с (8.15) функция  $\alpha(\Theta)$  описывает изменение величины измеренного градиента от ориентации яркостного перепада, причем  $\alpha(\Theta = 0) = 1$ . Как показывают результаты анализа, оператор Превитт-I обеспечивает очень высокую устойчивость обработки по отношению к ориентации идеального перепада яркости. При изменении угла  $\Theta$  от 0 до  $\pm 45^\circ$  отклонение  $\alpha(\Theta)$  от 1 не превышает  $\pm(0.04 \div 0.05)$ . Несколько больше диапазон изменений

$\alpha(\Theta)$  у оператора Собеля, у которого значения этой функции находятся в интервале  $[1...1.12]$ . Значительно хуже обстоит дело у оператора Робертса, поскольку при его применении при увеличении  $|\Theta|$  от  $0^\circ$  до  $45^\circ$  значение  $\alpha(\Theta)$  монотонно уменьшается от 1 до 0.4...0.45.

Второй характеристикой, также применяемой для оценки эффективности операторов, является чувствительность измеренного градиента к величине смещения  $d$  линии яркостного перепада относительно центра симметрии маски. Соответствующий показатель определяется выражением

$$\beta(d) = h(i, j|d) / h(i, j|d = 0), \quad (8.16)$$

смысл которого аналогичен (8.15). Результаты исследования этого показателя свидетельствуют о том, что с ростом  $d$  происходит быстрое уменьшение функции  $\beta(d)$  при всех видах операторов, описанных выше, причем практически отсутствует зависимость от ориентации перепада. При значениях  $d \geq 1.5$  (единицей измерения  $d$  является шаг, определяющий расстояние между отсчетами изображения)  $\beta(d) \approx 0$ . Быстрое уменьшение  $\beta(d)$ , происходящее при увеличении  $d$ , является положительной чертой масочных операторов, поскольку этим обеспечивается их высокая разрешающая способность.

Третьей характеристикой, применяемой для сравнения масочных операторов, является уровень выходных шумов. Предположим, что входное изображение содержит аддитивный белый шум с дисперсией  $\sigma^2$ , одинаковой во всех точках кадра. Легко найти дисперсию выходного шума  $\sigma_y^2$ , если учесть, что отклик масочного оператора представляет взвешенную сумму входных отсчетов, в которой в роли весовых коэффициентов выступают элементы  $a_{ij}$  матрицы  $\mathbf{A}$ . В результате имеем:

$$\sigma_y^2 = \sigma^2 \sum_{i=1}^3 \sum_{j=1}^3 a_{i,j}^2. \quad (8.17)$$

Помехоустойчивость масочного оператора определяют отношением сигнал/шум, получаемым на его выходе, поэтому дополнительно необходимо вычислить величину полезного эффекта. Для этого следует задаться моделью входного яркостного перепада. Полагая его ориентацию горизонтальной, считаем, что сиг-

нальная матрица (без учета шума), получаемая в рабочем окне, имеет вид:

$$S = \|s_{i,j}\| = 0.5s_0 \cdot \begin{vmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{vmatrix}.$$

При этом отношение “яркостный перепад/шум” для входного изображения характеризуется величиной  $q_{\text{вх}} = s_0/\sigma$ .

Полезный сигнал на выходе можно записать:

$$s_{\text{вых}} = \sum_{i=1}^3 \sum_{j=1}^3 s_{i,j} a_{i,j},$$

а отношение сигнал/шум после обработки

$$q_{\text{вых}} = \frac{\sum_{i=1}^3 \sum_{j=1}^3 s_{i,j} a_{i,j}}{\sigma \sqrt{\sum_{i=1}^3 \sum_{j=1}^3 a_{i,j}^2}}.$$

Результат применения этой формулы к различным операторам представлен в следующей таблице.

Значения  $q_{\text{вых}}$  для различных операторов

Превитт-1

Превитт-2

Собель

$$\sqrt{1.5}q_{\text{вых}}$$

$$\sqrt{0.75}q_{\text{вх}}$$

$$\sqrt{\frac{4}{3}}q_{\text{вх}}$$

Из таблицы видно, что максимальное отношение сигнал/шум обеспечивается оператором Превитт-1, наименьшую же его величину имеет оператор Превитт-2, оператор Собеля несколько уступает оператору Превитт-1. Причиной заметного проигрыша оператора Превитт-2 является упоминавшееся выше отклонение его матрицы  $A_{p2}$  (8.2) от канонического вида (8.14). Отличие от нуля элементов второй строки этой матрицы приводит, согласно (8.17), лишь к увеличению выходного шума, не внося при этом вклада, как видно из (8.13), в полезный выходной эффект.

Из таблицы также следует, что применение рационально построенных матриц подчеркивания контуров, в которых закладывается идея дифференцирования с накоплением, не только не уменьшает выходного отношения сигнал/шум, что свойственно

простым операторам, не использующим накопления, но даже, наоборот, приводит к его увеличению по сравнению с входным.

Таким образом, приведенные результаты говорят о высоком качестве операторов Превитт-1 и Собеля. Это, однако, нельзя рассматривать как окончательное решение проблемы подчеркивания контуров, а можно лишь как результат сравнительной оценки некоторых из известных процедур.

Получаемое на основе градиентной обработки (8.5) изображение  $h(i, j)$  обычно, как говорилось во введении, используют для образования контурного препарата — бинарного изображения, у которого контурные точки принимают значение 1 (уровень черного), а остальные — 0 (уровень белого) или наоборот. Бинаризацию выполняют путем сравнения значений  $h(i, j)$  с порогом  $h_0$  по правилу:

$$z(i, j) = \begin{cases} 1 & \text{при } h(i, j) \geq h_0 \\ 0 & \text{при } h(i, j) < h_0 \end{cases},$$

а порог можно выбирать автоматически на основе гистограммного анализа препарата  $h(i, j)$ , например, путем построения дискриминантной функции, применяемой при сегментации.

На рис. 8.1 приведены результаты эксперимента по градиентному подчеркиванию контуров и построению на этой основе контурного препарата. Исходное изображение, имеющее размер  $256 \times 256$  элементов, показано на рис. 8.1, а. Его функция яркости, соответствующая 220-й строке, приведена на рис. 8.1, б, а та же строка после обработки изображения оператором Превитт-1 для подчеркивания вертикальных перепадов — на рис. 8.1, в. Эта же строка после вычисления градиента показана на рис. 8.1, г. Рис. 8.1, д иллюстрирует вид поля градиентной функции  $h(i, j)$ , а рис. 8.1, е — вид контурного препарата, полученного при бинаризации градиента. Приведенные здесь результаты соответствуют оператору Превитт-1. Отметим в заключение, что визуальное отличие результатов обработки при применении оператора Собеля от приведенных несущественно.

## 8.2. РАНГОВОЕ ОБНАРУЖЕНИЕ ЛОКАЛЬНЫХ КОНТУРНЫХ ПРИЗНАКОВ

При выделении контуров происходит замена исходного полутонового изображения преобразованным, в котором исходные яркости заменены локальными контурными признаками (ЛКП). Если

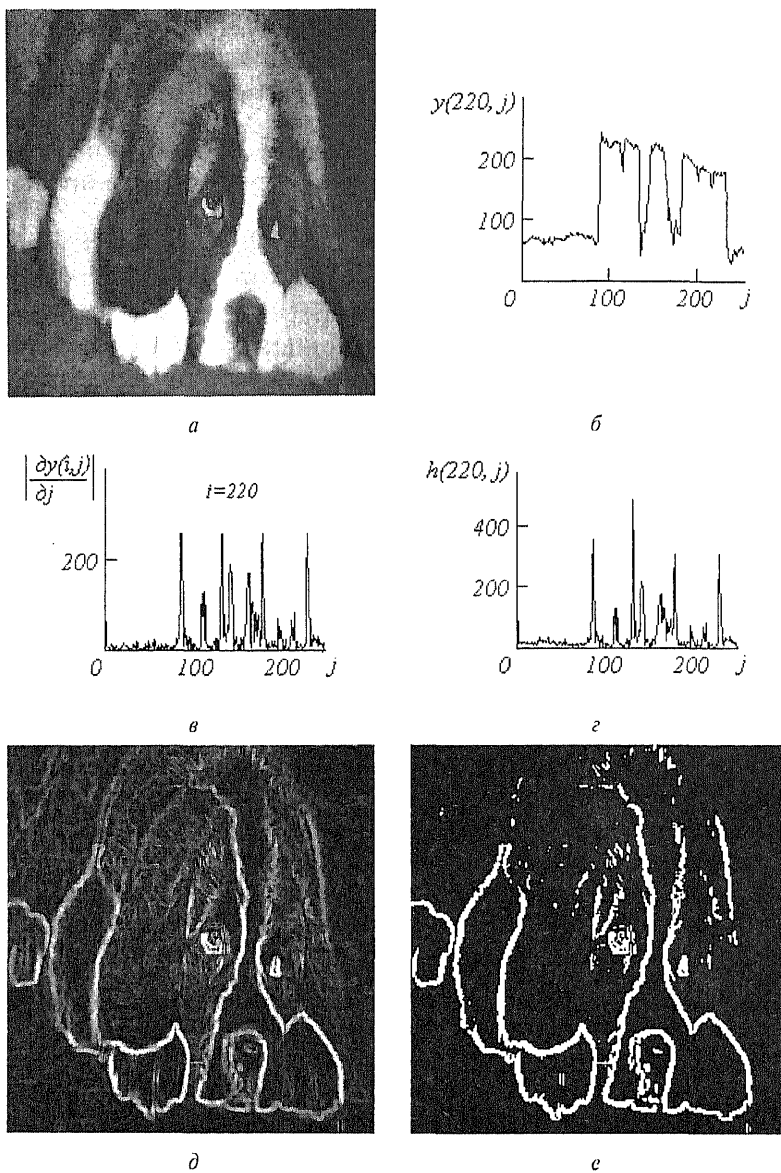


Рис. 8.1. Градиентное выделение контуров:

а – исходное изображение; б – функция яркости, 220-я строка; в – результат подчеркивания вертикальных перепадов, 220-я строка; г – градиент, 220-я строка; д – изображение градиента; е – контурный препарат

новое изображение, которое можно назвать контурным препаратом, далее используется для автоматической вторичной обработки (обнаружение или распознавание объектов, слежение за движущимся объектом, привязка изображений и др.), то к его характеристикам целесообразно предъявить определенные требования. Главное из них — это получение устойчивых результатов контурной обработки. Качество сформированного контурного препарата по возможности не должно зависеть от таких характеристик исходного изображения, как средняя яркость, динамический диапазон яркостей, неравномерная освещенность сцены и т. п. Рассмотренный выше подход к решению задачи не способен автоматически обеспечить эти качества, поскольку при изменении перечисленных характеристик изображения будут изменяться и результаты сравнения градиентного поля с порогом.

Весьма существенным является сочетание двух требований. Первое заключается в том, чтобы среднее количество ложно обнаруженных ЛКП было стабилизировано при вариации перечисленных выше неконтролируемых характеристик исходных изображений. В этом случае вторичная обработка находится всегда в состоянии фиксированной вычислительной нагрузки. Это позволяет рационально проектировать вычислительное устройство без неоправданных запасов его вычислительных ресурсов или при ограниченных ресурсах гарантирует выполнимость вычислений.

Второе требование состоит в необходимости обеспечить максимальную или близкую к ней вероятность правильного обнаружения истинных ЛКП.

Эти требования типичны для статистической проверки гипотез, применяемой при решении проблем обнаружения сигналов. В радиолокации получили мощное развитие методы обнаружения сигналов, которые сохраняют эффективность при априорно неизвестных характеристиках сигналов и помех. Вполне естественно, что эти методы начинают применяться и в задачах обработки изображений, поскольку условия их применения здесь достаточно близки к задачам радиолокационного обнаружения.

Одним из методов, применяемых в условиях априорной неопределенности при обработке изображений, является ранговое обнаружение сигналов. В данном разделе будет рассмотрено его применение к обнаружению локальных контурных признаков.

Существенная черта рангового обнаружения (РО) — его способность обеспечивать постоянство вероятности ложного обнаружения при независимой помехе. При обнаружении ЛКП в роли помехи выступает фон изображения, который для обычных сцен является случайной двумерной функцией с сильно корре-

лированными (следовательно, и зависимыми) значениями. Потому непосредственное применение РО к исходному изображению не оправдано. Обнаружению в этих условиях должна предшествовать предобработка, делающая отсчеты фона независимыми.

### 8.2.1. ДЕКОРРЕЛЯЦИЯ ФОНА ИЗОБРАЖЕНИЯ

Часто при вероятностном описании изображений достаточно точной является гауссовская модель. В этом случае независимость и некоррелированность элементов изображения эквивалентны друг другу. Поэтому предобработка сводится к такому преобразованию изображения, при котором его элементы становятся некоррелированными. Ограничимся далее случаем гауссовского распределения.

Для построения декоррелирующей процедуры воспользуемся широко известным описанием изображения при помощи стохастических разностных уравнений, применение которых уже рассматривалось и в настоящем пособии. Целесообразно использовать некаузальную модель, простейший вариант которой описывается уравнением:

$$\lambda(i, j) = \alpha_1[\lambda(i-1, j) + \lambda(i+1, j)] + \alpha_2[\lambda(i, j-1) + \lambda(i, j+1)] + \xi(i, j). \quad (8.18)$$

Здесь  $\lambda(i, j)$  — гауссовское поле, а  $\alpha_1, \alpha_2$  — параметры, характеризующие меру влияния соседних отсчетов на данный отсчет поля. Уравнение (8.18) описывает комбинированное действие двух механизмов.

Первые слагаемые в (8.18), заключенные в квадратные скобки, можно рассматривать как прогноз (предсказание) поля в точке  $i, j$  по соседним его отсчетам. В данном случае в прогнозе участвуют четыре соседние точки, конфигурация расположения которых в кадре соответствует используемой некаузальной модели.

Второй механизм представлен слагаемым  $\xi(i, j)$ , также являющимся случайным полем с гауссовским законом распределения. С его помощью описывается шум или ошибка предсказания. Если первые слагаемые характеризуют передачу

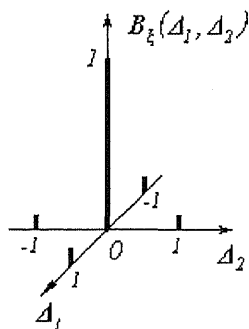


Рис. 8.2. Корреляционная функция шума предсказания



информации окружающих точек в точку  $i, j$ , то последнее описывает *новую информацию*, содержащуюся в данной точке.

Некоторые результаты изучения моделей случайных полей вида (8.18), полученные Д. Вудсом, приведены в работах [8.2, 8.3]. Доказано, что, если фон  $\lambda(i, j)$  является марковским полем, то корреляционная функция  $B_{\xi}(\Delta_1, \Delta_2)$  шума предсказания  $\xi(\cdot)$  имеет вид, качественно показанный на рис. 8.2. Ее важной особенностью является равенство нулю при всех значениях  $\Delta_1, \Delta_2$ , за исключением четырех точек при  $\Delta_1 = \pm 1$  или  $\Delta_2 = \pm 1$ , в которых  $B_{\xi}(\cdot) = -\alpha_1$  или  $B_{\xi}(\cdot) = -\alpha_2$  соответственно. При этом достоинство модели (8.18) состоит и в том, что описываемое с ее помощью случайное поле очень близко к изотропному (при  $\alpha_1 = \alpha_2$ ), а изотропность характерна для большинства реальных изображений.

С учетом этих свойств модели фона (8.18) воспользуемся для его декорреляции процедурой:

$$\begin{aligned} \tilde{\lambda}(i, j) = \lambda(i, j) - \alpha_1[\lambda(i-1, j) + \lambda(i+1, j)] - \\ - \alpha_2[\lambda(i, j-1) + \lambda(i, j+1)] = \xi(i, j). \end{aligned} \quad (8.19)$$

Здесь и в дальнейшем символом « $\sim$ » будем обозначать продукт декорреляции. Как видно из (8.19), результат этой обработки совпадает с шумом предсказания  $\xi(\cdot)$ , который по свойствам своей корреляционной функции близок к белому шуму. Для того чтобы при обнаружении ЛКП использовать независимую выборку, ее отсчеты следует располагать так, чтобы их взаимная корреляция равнялась нулю. При этом следует руководствоваться видом корреляционной функции, изображенной на рис 8.2. Если же  $\alpha_1 \ll 1$ ,  $\alpha_2 \ll 1$ , то в качестве приближенно независимой выборки можно использовать все элементы декоррелированного изображения без геометрических ограничений.

Процедура декорреляции воздействует и на сигнальную составляющую изображения. Необходимо оценить изменение яркостного перепада, происходящее при этом воздействии. Представим полную модель исходного изображения  $y(i, j)$  в виде суммы непрерывной (фоновой) составляющей  $\lambda(i, j)$  и ступенчатой функции  $s(i, j)$ , описывающей яркостный перепад:

$$y(i, j) = \lambda(i, j) + s(i, j). \quad (8.20)$$

Полагаем, что функция  $s(\cdot)$  принимает одно из двух значений  $s_0$  или  $s_1 > s_0$ , причем в пределах локальных участков граница, разделяющая области малой и большой яркости, прямолинейна. Считается, что линия границы может иметь одну из четырех ориентаций — вдоль координатных направлений и под углами  $45^\circ$  к ним. В результате воздействия декоррелирующей процедуры на изображение  $y(i, j)$  формируется преобразованное изображение:

$$\begin{aligned}\tilde{y}(i, j) &= y(i, j) - \alpha_1[y(i-1, j) + y(i+1, j)] - \\ &\quad - \alpha_2[y(i, j-1) + y(i, j+1)] = \\ &= \tilde{\lambda}(i, j) + \tilde{s}(i, j) = \xi(i, j) + \tilde{s}(i, j),\end{aligned}$$

состоящее из суммы преобразованных сигнала  $\tilde{s}(i, j)$  и фона  $\tilde{\lambda}(i, j) = \xi(i, j)$ . Предположим, что сигнальная функция  $s(i, j)$  представляет собой яркостный перепад, ориентированный вдоль столбцов изображения, что символически показано в верхней части рис. 8.3. Полагаем, что скачок яркости происходит на  $j$ -м столбце, т.е. исходный сигнал описывается выражением:

$$s(i, j_1) = \begin{cases} s_0 & \text{при } j_1 < j, \\ s_1 & \text{при } j_1 \geq j. \end{cases}$$

Вид сигнала, прошедшего через процедуру декорреляции, зависит от значений параметров  $\alpha_1, \alpha_2$ , поэтому остановимся на их определении. Примем условие одинаковости характеристик изображения по строкам и столбцам, при котором  $\alpha_1 = \alpha_2 = \alpha$ . Поскольку параметр  $\alpha$  участвует в процедуре некаузального предсказания (8.18), его выбор необходимо подчинить минимизации ошибок этого предсказания:

$$\begin{aligned}E\{\lambda(i, j) - \alpha[\lambda(i-1, j) + \lambda(i+1, j) + \lambda(i, j-1) + \lambda(i, j+1)]\}^2 = \\ = \min_{\alpha}.\end{aligned}$$

Приравнивая нулю производную по  $\alpha$  левой части этого выражения, находим:

$$\alpha = \frac{R_{1,0} + R_{0,1}}{(1 + 2R_{1,1} + R_{2,0}) + (1 + 2R_{1,1} + R_{0,2})}, \quad (8.21)$$

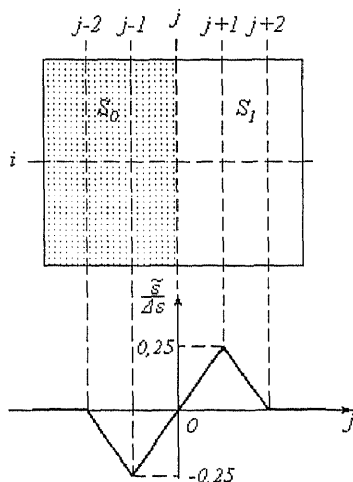


Рис. 8.3. Преобразование яркостного перепада при декорреляции

где  $R_{\Delta_1, \Delta_2}$  — нормированная корреляционная функция фона, вид которой необходимо далее конкретизировать. Большинство реальных изображений обладает свойством изотропности, при котором корреляционные свойства одинаковы по всем направлениям. Этому отвечает часто применяемая аппроксимация вида:

$$R_{\Delta_1, \Delta_2} = \rho \sqrt{\Delta_1^2 + \Delta_2^2}, \quad (8.22)$$

параметр  $0 < \rho < 1$  которой имеет смысл одношагового коэффициента корреляции, поскольку  $R_{1,0} = R_{0,1} = \rho$ . Подставляя данную аппроксимацию в (8.21), находим

$$\alpha = \frac{\rho}{1 + 2\rho\sqrt{2} + \rho^2}.$$

Очень часто значения коэффициента одношаговой корреляции  $\rho$  лежат в пределах  $0,9 < \rho < 1$ . Этому соответствует  $0,25 < \alpha < 0,255$ . Изменение параметра  $\alpha$  в таких небольших пределах незначительно изменяет отклик оператора декорреляции на яркостный перепад  $s(i, j)$ . Поэтому в нижней части рис. 8.3 приведено сечение выходного сигнала, рассчитанное при  $\alpha = 0,25$ . Здесь  $\Delta s = s_1 - s_0$  — абсолютная величина яркостного перепада. Из рисунка видно, что преобразованный сигнал

имеет два односточечных пика одинаковой величины и противоположных знаков.

Отметим, что приведенный пример соответствует входному перепаду типа "от темного к светлому". Характер отклика остается таким же и при противоположном перепаде типа "от светлого к темному". В этом случае лишь изменяется очередность следования импульсов положительной и отрицательной полярности.

### 8.2.2. РАНГОВОЕ ОБНАРУЖЕНИЕ ЛКП

Для обнаружения ЛКП воспользуемся двухвыборочным алгоритмом Вилкоксона [8.4]. При этом формируем две выборки, одна из которых  $X = [x_1, \dots, x_n]^T$  является опорной и принадлежит области фона, другая  $Y = [y_1, \dots, y_m]^T$  — рабочей. В нее включаем элементы изображения, относительно которых решается вопрос о принадлежности к контурным точкам. При этом входящие в  $Y$  элементы могут быть как точками фона, так и контурными точками изображения. Способы получения этих выборок могут быть различными. Необходимо лишь обеспечивать их независимость, а при формировании рабочей выборки руководствоваться, кроме того, видом геометрической структуры яркостного перепада после его преобразования декоррелирующим оператором. Другими словами, в состав вектора  $Y$  следует включать те элементы преобразованного изображения, которые соответствуют ожидаемому положению положительных и отрицательных сигнальных выбросов.

Рассмотрим сначала построение рангового обнаружителя в более простых условиях, когда все сигнальные импульсы имеют положительную полярность. Для каждого элемента  $y_i$  рабочего вектора  $Y$  и всех элементов опорной выборки  $X$  составляем *вариационный ряд*, в котором входящие в него величины располагаются в порядке возрастания. Номер  $r_i$ , определяющий число опорных элементов, предшествующих рабочему элементу  $y_i$  в этом ряде, называется его *рангом*. Далее определяем *ранговую статистику*

$$R = \sum_{i=1}^m r_i, \quad (8.23)$$

сравнение которой с порогом  $R_0$  приводит к решению о наличии или отсутствии ЛКП согласно правилу:

$$R \begin{cases} \geq R_0 & \text{ЛКП имеется} \\ < R_0 & \text{ЛКП отсутствует.} \end{cases} \quad (8.24)$$

При фактическом отсутствии ЛКП элементы обоих векторов  $X$  и  $Y$  образуют однородное множество случайных величин, подчиняющихся одному и тому же закону распределения  $w(x)$ . В этом случае элемент рабочей выборки  $y_i$  с равными шансами занимает любое положение в вариационном ряду, а это означает, что случайная величина  $r_i$  подчиняется равновероятному закону распределения:

$$p(r_i) = \frac{1}{n+1}, \quad r_i = 0, 1, 2, \dots, n.$$

Подчеркнем, что условием равновероятности является независимость и однородность выборки. Что особенно важно, на данное распределение не оказывает влияния конкретный вид распределения декоррелированного фона  $w(x)$ .

Но тогда и закон распределения  $w(R)$  решающей статистики  $R$  при фактическом отсутствии в  $Y$  яркостного перепада не зависит от распределения фона  $w(x)$ . Следовательно, от  $w(x)$  не зависит и вероятность ложного обнаружения ЛКП, которому соответствует верхнее условие в процедуре (8.24). Для практического применения особенно важно, что и выбор порога ранговой процедуры  $R_0$ , который подчиняют требованию получения желаемой величины вероятности ложного обнаружения, также не связан с распределением фона  $w(x)$ .

Таким образом, применение рангового алгоритма автоматически стабилизирует вероятность ложного обнаружения, делая ее независимой от вида (и тем более от параметров) закона распределения фона. Подчеркнем, что это свойство достигается только при условии, что все случайные величины, образующие выборки  $X$  и  $Y$ , независимы, поскольку в противном случае в закон распределения  $w(R)$  будут входить параметры, описывающие зависимость элементов.

При наличии полезного сигнала элементы вектора  $Y$  статистически превышают элементы  $X$ , поэтому закон распределения ранга  $p(r_i)$  перестает быть равномерным: вероятности больших значений  $r_i$  увеличиваются за счет уменьшения вероятностей малых значений. Эта тенденция проявляется тем сильнее, чем

больше величина яркостного перепада  $\Delta s$ . Тем больше становится и вероятность выполнения верхнего неравенства в (8.24), которая теперь является вероятностью правильного обнаружения.

Для расчета характеристик обнаружения необходимо иметь закон распределения  $w(R)$  решающей статистики  $R$  как при отсутствии, так и при наличии яркостного перепада. При достаточно больших значениях  $n$  и  $m$  распределение  $w(R)$  в соответствии с центральной предельной теоремой теории вероятностей стремится к гауссовскому, полностью определяемому своими математическим ожиданием и дисперсией. Реально гауссовским приближением можно пользоваться при  $m > 3$  и  $n > 20$  [8.4]. Однако даже нахождение моментов распределения  $w(R)$  удастся выполнить только при отсутствии полезного сигнала. Вместе с тем именно это является наиболее важным, поскольку знание  $w(R)$  при отсутствии сигнала дает возможность определения порога  $R_0$  по заданному значению вероятности ложного обнаружения  $F$ . В работе [8.4] показано, что с учетом гауссовского приближения порог находится по формуле

$$R_0 = R_F \left[ nm(1+n+m)/12 \right]^{1/2} + nm/2, \quad (8.25)$$

где  $R_F$  – процентная точка гауссовского распределения, соответствующая вероятности  $F$ . Как и следовало ожидать, величина порога зависит только от объемов рабочей и опорной выборок и вероятности ложного обнаружения.

Вернемся теперь к обнаружению ЛКП с учетом биполярного характера полезного сигнала, возникающего после воздействия на него процедуры декорреляции. Отрицательные значения половины всех сигнальных выбросов не будут способствовать отмеченному выше перераспределению вероятностей  $p(r_i)$  для этих элементов. Более того, вероятности больших значений ранга будут, наоборот, уменьшаться за счет роста вероятностей малых значений. Таким образом, для сигнальных элементов  $y_i$ , содержащих отрицательные выбросы, наблюдается тенденция, противоположная той, которая имеет место для положительных выбросов и на которой основано действие обнаружителя. Для исключения этого явления необходимо перед процедурой ранжирования ожидаемые элементы сигнала с отрицательными значениями умножать на коэффициент  $-1$ . Очевидно, это снимает отмеченный недостаток описанной процедуры. Вместе с тем при

определенных условиях не возникает каких-либо проблем в работе обнаружителя при отсутствии в рабочей выборке  $Y$  полезного сигнала, т.е. обеспечивается стабильность вероятности ложного обнаружения  $F$  при выборе порога в соответствии с (8.25). Условие, при котором сохраняется непараметрическое свойство обнаружителя, состоит в симметрии плотности распределения  $w(x)$  фона. Действительно, в этом случае умножение некоторых элементов выборки на  $-1$  не изменяет ее однородности.

Представление об эффективности обнаружителя ЛКП дают *характеристики обнаружения*. Как и в радиолокации — это зависимости вероятности правильного обнаружения ЛКП  $D$  от относительной величины яркостного перепада  $q^2 = (\Delta s)^2 / \sigma_\lambda^2$ , где  $\sigma_\lambda^2$  — дисперсия исходного (коррелированного) фона. Пример характеристик обнаружения приведен на рис. 8.4. В качестве параметра в каждое из семейств входит вероятность ложного обнаружения ЛКП  $F$ .

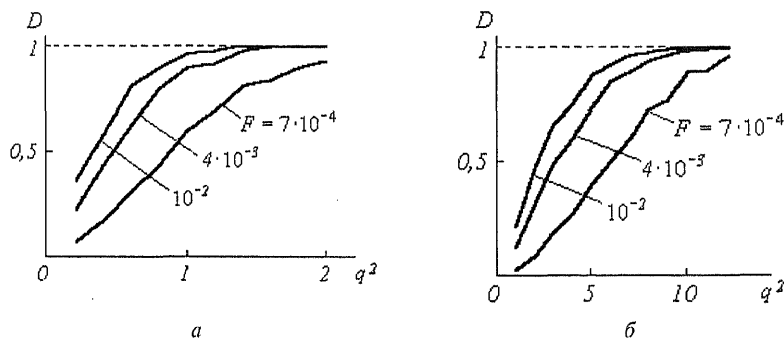


Рис. 8.4. Характеристики рангового обнаружения яркостных перепадов:  
 а — диагональное расположение ЛКП;  
 б — расположение ЛКП вдоль строк  
 или столбцов

Как уже говорилось, вероятность ложного обнаружения определяется расчетным путем в гауссовском приближении для распределения ранговой статистики  $w(R)$ . Определение же моментов этого распределения *при наличии* ЛКП представляет сложную задачу. Трудность связана с тем, что в этом случае выборка, участвующая в процедуре, не является однородной, поскольку в нее входят как опорные, так и рабочие элементы. Поэтому для нахождения вероятности правильного обнаружения  $D$  приходится использовать метод статистического моделирования. При нахож-

дении приведенных на рис. 8.4 характеристик обнаружения выполнялось моделирование яркостного перепада, наблюдавшегося на гауссовском фоне с изотропной корреляционной функцией вида (8.22) при коэффициенте одношаговой корреляции  $\rho = 0,9$ . Объем выборки для нахождения каждого значения  $D$  составлял 500. Несколько изрезанный характер зависимостей является признаком их статистической природы и не вполне достаточного объема статистического эксперимента.

Приведенные зависимости соответствуют двум ориентациям яркостного перепада — вдоль диагоналей (рис. 8.4, *а*) и вдоль координатных направлений (рис. 8.4, *б*). Из их сравнения видно, что эффективность обнаружения диагональных перепадов заметно выше, чем расположенных вдоль строк или столбцов. Дело в том, что оператор декорреляции по разному воздействует на яркостные перепады, имеющие различное расположение относительно координатных осей. Это, конечно, его недостаток, объясняемый не вполне изотропным характером его механизма. К сожалению, в настоящее время не найдены декоррелирующие операторы, обладающие полностью изотропными характеристиками.

Результаты эксперимента, иллюстрирующего ранговое обнаружение ЛКП, показаны на рис. 8.5. Здесь использована модель объекта в виде квадрата, средняя яркость  $s_1 = 118$  которого выше, чем средняя яркость фона  $s_0 = 100$ . Флюктуации фона моделировались гауссовским полем, имеющим корреляционную функцию (8.22) с коэффициентом одношаговой корреляции  $\rho = 0,9$  и дисперсией  $\sigma_\lambda^2 = 50$ . Нетрудно убедиться, что при этом  $q^2 = 7$ .

На рис. 8.5, *б* и 8.5, *в* приведена исходная картина, на которую в виде белых точек нанесены результаты срабатывания рангового обнаружителя при двух значениях вероятности ложного обнаружения. Видно, что снижение вероятности ложного срабатывания приводит к пропорциональному уменьшению числа ложно отмеченных белых точек. Более детальный анализ результатов эксперимента показывает, что количество ложных отметок соответствует теоретически ожидаемому значению этой величины.

Вторая серия экспериментов, представленная на рис. 8.6, позволяет сравнить выделение контурных линий оператором Превитта и ранговым обнаружителем. Исходная картина, показанная на рис. 8.6, *а*, как и в предыдущем эксперименте, содержит изображение объекта в виде светлого квадрата, наблюдаемого на гауссовском коррелированном изотропном фоне.



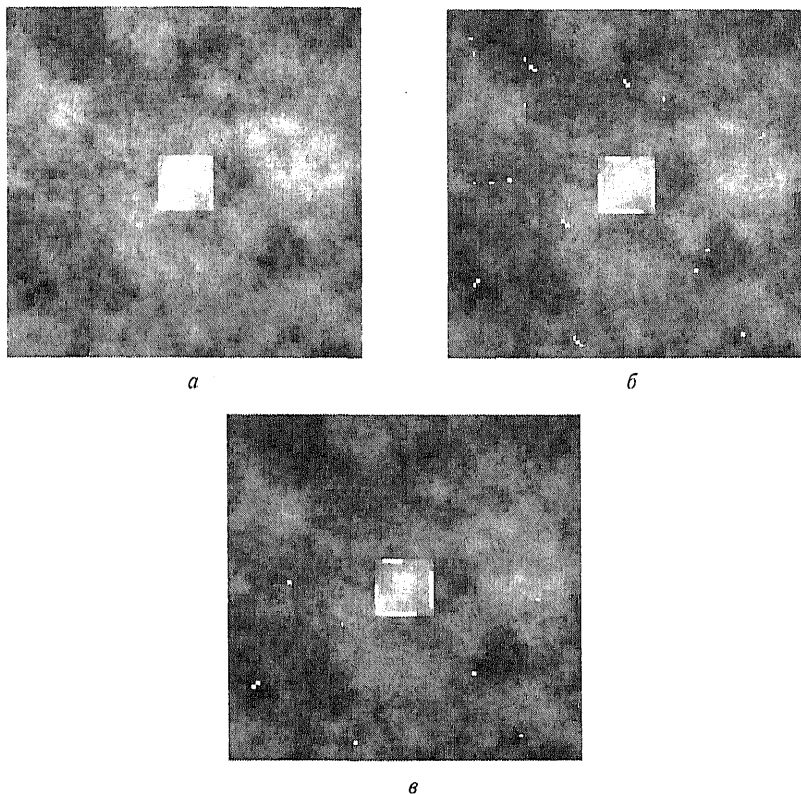


Рис. 8.5. Ранговое обнаружение ЛКП:

*a* — исходное изображение; *б* — результат обнаружения; при  $F = 10^{-3}$ ; *в* — результат обнаружения при  $F = 5 \cdot 10^{-4}$

Особенность же этой картины состоит в наличии сильного модулирующего пятна, действие которого поражает и область расположения объекта. На рис. 8.6, *б* приведен бинаризованный результат действия оператора Превитта. Порог бинаризации подбирался таким образом, чтобы при относительно неплохой передаче контура объекта (для этого порог должен быть не слишком высоким) количество ложных срабатываний было не очень большим (это, наоборот, требует повышения порога). Результат действия рангового оператора обнаружения представлен на рис. 8.6, *в* при вероятности ложного срабатывания  $F = 0.0007$ .

Оператор Превитта требует специального участия наблюдателя для подбора порога. При этом не гарантируется заданная сте-

пень засорения полученного в результате препарата. Ранговый обнаружитель является полностью автоматическим. Качество выделения объекта во втором случае оказывается бесспорно более высоким, обнаружитель проявляет способность “заглянуть” под модулирующее пятно.

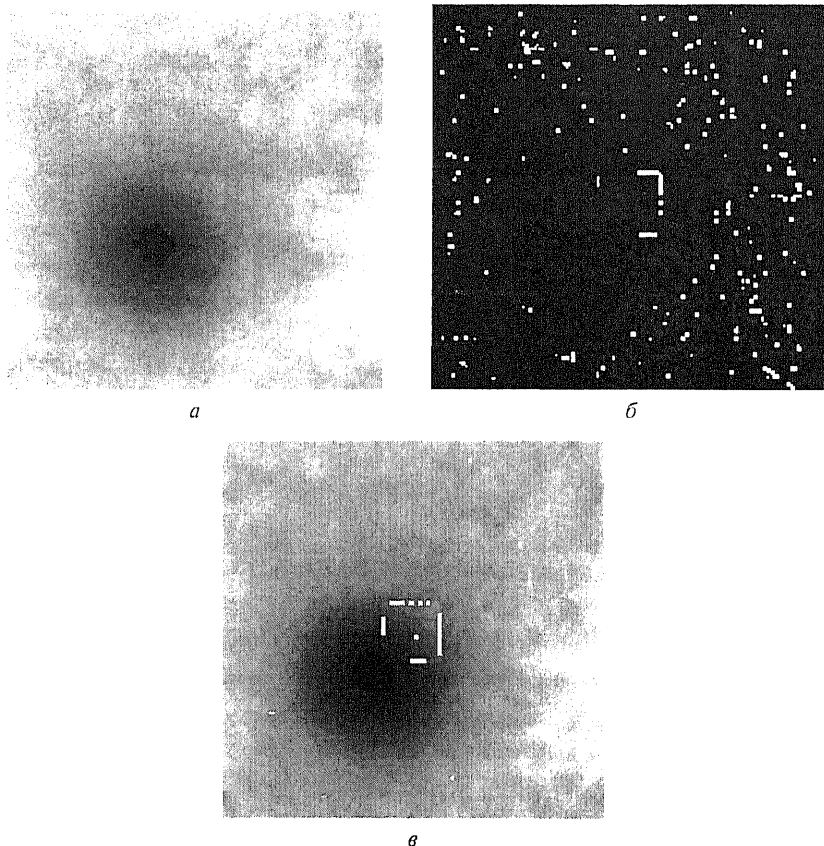


Рис. 8.6. Выделение контуров на неоднородном изображении:  
*а* — исходное изображение; *б* — выделение контура оператором Превитта;  
*в* — выделение контура ранговым обнаружителем

Отметим в заключение, что описанный в данном разделе подход к выделению ЛКП не следует рассматривать как завершённый безукоризненный вариант контурной обработки изображений. В большей степени он иллюстрирует необходимость и целесообразность применения к решению данной задачи методов, известных из смежных наук.

**ВОПРОСЫ К ГЛАВЕ 8**

8.1. Поясните принцип формирования матриц для операторов подчеркивания контуров, выполняющих сглаженное дифференцирование в горизонтальном и диагональных направлениях.

8.2. Поясните смысл выражений (8.5), (8.6), определяющих величину и направление градиента яркости.

8.3. Какими характеристиками оценивают качество работы градиентных операторов подчеркивания контуров?

8.4. Какие требования предъявляются к функции  $\alpha(\Theta)$ , описывающей чувствительность градиентных операторов к ориентации яркостного перепада?

8.5. Какие требования предъявляются к функции  $\beta(d)$ , описывающей чувствительность градиентных операторов к относительному смещению перепада и центра симметрии окна?

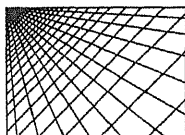
8.6. Каковы основные достоинства использования рангового обнаружения для выделения контуров?

8.7. Для чего и как выполняется декорреляция фона изображения? Какой вид имеет корреляционная функция фона после его декорреляции?

8.8. Что происходит с функцией яркостного перепада при декорреляции и какие последствия это имеет для рангового обнаружения?

8.9. Поясните механизм действия рангового обнаружителя яркостного перепада.

8.10. При каких условиях и почему ранговый обнаружитель стабилизирует вероятность ложного обнаружения яркостных перепадов?



## ГЛАВА 9

### РАСПОЗНАВАНИЕ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

#### 9.1. ОСНОВНЫЕ ПОЛОЖЕНИЯ

Обнаружение и распознавание объектов составляет неотъемлемую часть человеческой деятельности. Пока еще не совсем понятно, как человеку удастся так точно и так быстро выделять и узнавать нужные предметы в разнообразии окружающей среды. Попытки выяснить это делаются физиологами и психологами уже более ста лет. Однако здесь наша цель состоит не в понимании механизма восприятия человека (и животных), а в описании методов автоматизированного распознавания объектов по их изображениям: новой информационной технологии, мощной, практичной и в некотором смысле универсальной методологии обработки и оценивания информации и выявления скрытых закономерностей [9.1].

Распознавание трехмерных объектов по их двумерным изображениям стало в последнее время одной из важнейших задач анализа сцен и машинного зрения. Исходную для распознавания информацию содержат изображения в различных частях полного спектра излучений (оптические, инфракрасные, ультразвуковые и т.д.), полученные различными способами (телевизионные, фотографические, лазерные, радиолокационные, радиационные и т.д.), преобразованные в цифровую форму и представленные в виде некоторой числовой матрицы. Под объектом понимаем не только (и не столько) цифровое представление локального фрагмента двумерной сцены, а некоторое его приближенное описание, в виде набора характерных свойств (*признаков*). Основное назначение описаний (образов объектов) — это их использование

в процессе установления соответствия объектов, осуществляемого путем сравнения (сопоставления). Задачей распознавания является определение «скрытой» принадлежности объекта к тому или иному классу путем анализа вектора значений наблюдаемых признаков. Информацию о связи между значениями признаков объекта и его принадлежностью к определенному классу алгоритм распознавания должен извлечь из обучающей совокупности объектов, для которых известны либо значения и признаков и классов, либо только значения их признаков. В первом случае задача называется задачей обучения распознаванию образов с учителем, а во втором — без учителя. Здесь предполагается что каждый объект «принадлежит» одному образу из некоторого фиксированного множества. При отнесении (классификации) объектов требуется применить некоторое установленное ранее правило, чтобы решить, какому образу (классу) принадлежит объект. В задаче *расознавания с обучением* правило классификации должно вырабатываться на основе исследования множества объектов с известной принадлежностью различным классам. Эти объекты в совокупности называются *обучающим множеством* или *выборкой*. В задаче *автоматического формирования образов* объекты предъявляются «наблюдателю» без указания их принадлежности классам (расознавание без учителя). Наблюдатель (алгоритм распознавания) должен самостоятельно построить соответствующее определение классов (кластерный анализ). Разумеется, такой подход к анализу изображений адекватен лишь одному из двух аспектов двуединой задачи обнаружения и распознавания объектов сцены, а именно, собственно распознаванию класса вполне определенного (выделенного) фрагмента изображения, рассматриваемого как внешнее проявление некоторого скрытого образа. При этом вынужденно предполагается уже решенной *задача сегментации*, т. е. определение границ фрагментов, каждый из которых допустимо рассматривать как единое целое (объект).

Исследования по распознаванию образов пространственных объектов отличаются большим разнообразием в постановке задач и выборе средств их решения (методов обработки соответствующих фрагментов изображений), что является следствием разнообразия областей практического применения. Традиционными задачами, решавшимися еще в первых опытных разработках систем машинного зрения, служат задачи обнаружения и распознавания объектов, имеющих заданную форму на основе зашумленных и (возможно) деформированных изображений. Так, одной из первых практических задач, стимулировавших становление и

развитие теории распознавания объектов, была задача идентификации и распознавания человеческих лиц.

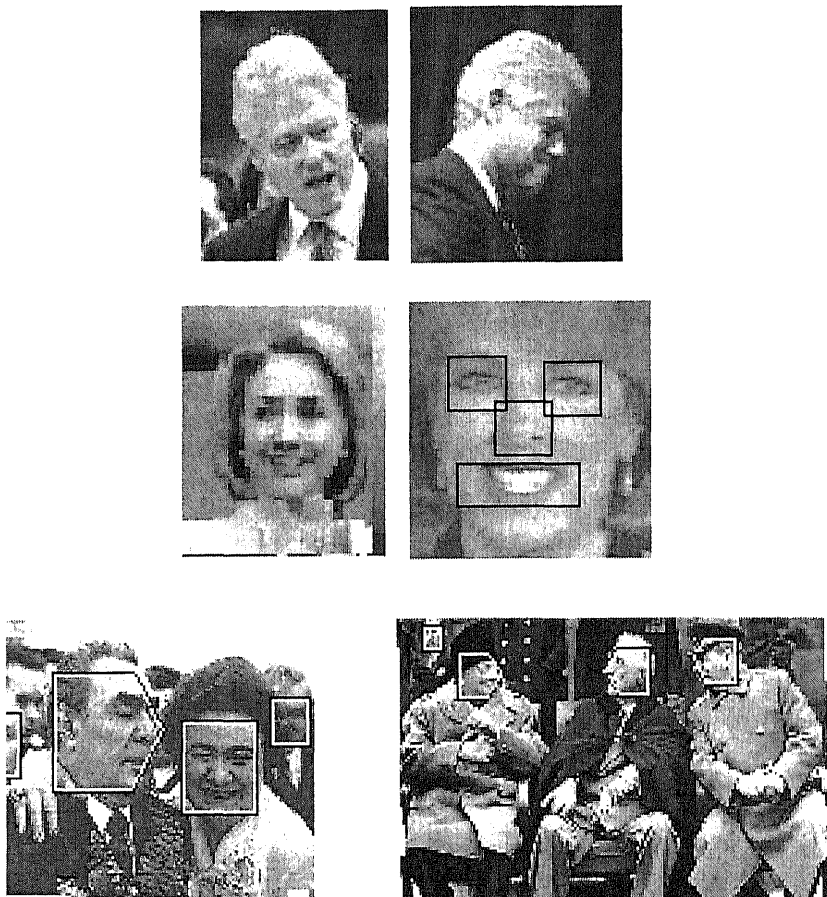


Рис. 9.1. Иллюстрация проблемы распознавания человеческих лиц и подходов к выбору информативных фрагментов (источник [9.15])

Сложность этой задачи обусловлена многообразием возможных ракурсов (масштабов, положений, углов поворота) распознаваемых лиц (см. рис. 9.1). Здесь предварительно необходимо построить внутреннее представление объектов, включающее проекции изображений [9.2, 9.15]. Данная задача до сих пор имеет широкое применение в системах охраны, при верификации кредитных карточек, в криминалистической экспертизе, на теле-

конференциях и т.д. Для ее решения предложены методы распознавания, основанные на теории нейрокомпьютерных сетей, корреляционно-экстремальных алгоритмах, методах вычисления статистических и алгебраических моментов, контурном анализе, 3D-моделировании и др. [9.2, 9.3]. Среди них особое внимание уделяется направлению, связанному с автоматическим выделением характерных (информативных) признаков объектов сцены, в данном случае элементов глаз, носа, рта, подбородка — рис. 9.1.

Немного позже, в задачах мониторинга (контроля) природной среды по аэрокосмическим изображениям наметилось еще несколько важных подходов к построению информативных признаков. В частности, когда были получены первые многозональные и спектрозональные данные дистанционного зондирования, большинство разработчиков машинных методов интерпретации обратились к изучению спектральных свойств природных объектов, поскольку яркости соответствующих им элементов изображений в различных спектральных диапазонах позволяли идентифицировать их при относительно низких вычислительных затратах [9.4].

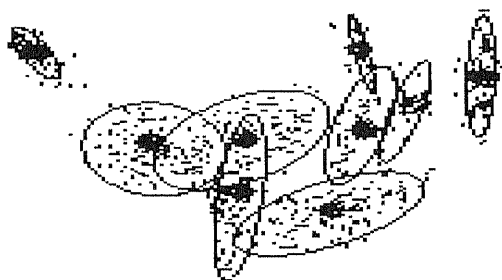


Рис. 9.2. Кластер-анализ спектральной яркости природных объектов, отображенный в двумерную плоскость пары информативных спектральных каналов

Наиболее употребительным методом обработки была «классификация без учителя» на основе кластерного анализа, с помощью которого пространство спектральных признаков разбивалось на различные группы (кластеры, см. рис. 9.2), а классификация элементов изображений позволяла одновременно сегментировать сцену на спектрально однородные области [9.5, 9.6].

Кроме того, выяснилось, что при распознавании природных образований помимо спектрозональных признаков оказались чрезвычайно важными также признаки, характеризующие локальную пространственную организацию полутонов (текстуру) объектов анализа. Опытный оператор-дешифровщик полагается на такую информацию (форму, ориентацию, распределение характерных элементов, контекст и другие пространственные характеристики) иногда в большей степени, чем на спектрозональные признаки [9.4]. В любом случае привлечение текстурных признаков позволяет существенно повысить надежность распознавания и увеличить количество различимых классов природных объектов.

В экспериментальных исследованиях, ориентированных на решение конкретных задач, исходными данными является некоторое множество фрагментов изображений (объектов сцены), дешифрованное специалистами-предметниками и служащее обучающим и контрольным материалом. Здесь первичная цель разработки алгоритма распознавания заключается в получении ответа на вопрос, возможен ли автоматизированный анализ и классификация соответствующих изображений объектов и, если да, то какой набор дешифровочных признаков наиболее эффективен и какую структуру должны иметь решающее правило и метод дешифрования.

## 9.2. РАЗДЕЛЯЮЩИЕ ФУНКЦИИ

Будем предполагать, что существует известное множество  $\tilde{C} = \{C_i\}$ , состоящее из  $k$  взаимно исключающих классов (образов) объектов. Каждый объект представляем набором результатов измерений выбранных признаков  $(x_1, x_2, \dots, x_m)^t = \mathbf{x}$ , называемых его описанием. Таким образом, описание объекта — это точка (вектор)  $\mathbf{x}$  в  $m$ -мерном пространстве признаков  $\mathfrak{R}^m$ . Евклидово расстояние между двумя точками  $\mathbf{x}$  и  $\mathbf{y}$  в  $\mathfrak{R}^m$

$$r(\mathbf{x}, \mathbf{y}) = \left( \sum_{i=1}^m (x_i - y_i)^2 \right)^{1/2}. \quad (9.1)$$

*Процедурой распознавания образов* объектов называется процедура определения областей  $\{R_i\}$  пространства  $\mathfrak{R}^m$  путем исследования описаний заданного множества  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$  объектов, про которые известно, каким классам они на самом деле



принадлежат. Иначе говоря,  $X$  будет обучающей выборкой, с которой работает процедура распознавания образов объектов.

*Процедурой классификации объектов* является процедура, отно-  
сящая объект к классу  $C_i$  тогда и только тогда, когда его описа-  
ние  $x$  попадает в область  $R_i$  пространства  $\mathbb{R}^m$ , соответствующую этому классу. Такая процедура классификации корректна, если объект и в самом деле относится к классу  $C_i$ . Таким образом, процесс распознавания объектов включает процедуру распознавания образов (формирование классов объектов) и процедуру классификации объектов (правило отнесения объектов к тому или иному классу). Геометрические понятия, включенные в распознавание объектов, достаточно просты. Рассмотрим, например, двумерное пространство признаков (рис. 9.3). Здесь представлены два класса  $\{C_1, C_2\}$  и обучающая выборка объектов. Проблема состоит в классификации неизвестного объекта — отнесении его к одному из классов. Одним из множества возможных решений является определение *разделяющей границы*, такой, например, как указана на рис. 9.3. Наше *решающее правило* тогда сводится к отнесению неизвестного объекта к классу  $C_1$ , если соответствующая ему точка (вектор признаков) лежит с одной стороны границы и к классу  $C_2$  — с другой. В данном случае неизвестный объект будет отнесен к классу  $C_2$ .

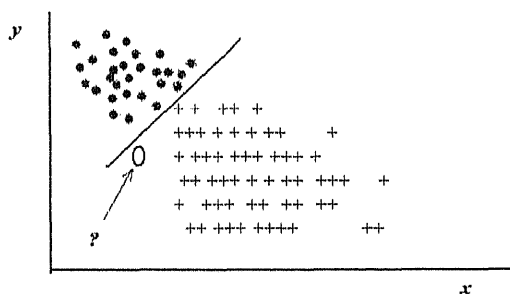


Рис. 9.3. Двумерное пространство признаков с обучающей выборкой: темными точками помечены объекты класса  $C_1$ , крестиками — объекты класса  $C_2$ ; эллипс — неизвестный объект

Из представленного примера видно, что две совокупности объектов оказалось удобно разделить прямой. Пусть

$d(x) = w_1x_1 + w_2x_2 + w_3 = 0$  — уравнение разделяющей прямой, где  $w_i$  — параметры, а  $x_i$  — признаки объектов. Функцию  $d(x)$  можно использовать в качестве *разделяющей* (дискриминантной) функции, поскольку, рассматривая объект  $x$ , принадлежность которого неизвестна, можно утверждать, что он принадлежит классу (образу)  $C_1$ , если  $d(x) > 0$ , и классу  $C_2$ , если  $d(x) < 0$ . Если вектор признаков  $x$  лежит на разделяющей границе, то имеет место неопределенность:  $d(x) = 0$ . Очевидно, что данный метод можно распространить на большее число классов и на более общий случай нелинейных границ в любом конечномерном евклидовом пространстве признаков [9.7, гл.2].

В общем случае пространство признаков можно считать состоящим из  $k$  областей  $R_i$ ,  $i = 1, \dots, k$ , каждая из которых содержит точки, соответствующие объектам из одного класса. При этом задача распознавания может рассматриваться как построение границ областей решений, разделяющих  $k$  классов, исходя из зарегистрированного (обучающего) множества векторов-признаков. Пусть эти границы определяются, например, разделяющими функциями  $d_1(x)$ ,  $d_2(x)$ , ...,  $d_k(x)$ . Если  $d_j(x) > d_i(x)$  для всех  $j \neq i$ , то объект  $x$  принадлежит классу  $C_i$ . Отсюда становится понятным, что основная проблема, возникающая после определения разделяющих функций, состоит в отыскании граничных поверхностей. Для их определения и нужна обучающая выборка объектов. Здесь процесс обучения считается законченным, когда проведены все разделительные границы между множествами образов (классов). Это удастся обычно выполнить при условии так называемой *компактности*, когда каждая группа (кластер) векторов-признаков, принадлежащих одному классу, образует в пространстве описания компактную (локально ограниченную) в некотором смысле область. Если кластеры, соответствующие различным классам, разнесены достаточно далеко друг от друга, то с успехом можно воспользоваться простыми схемами распознавания, например такими, как классификация объекта по расстоянию от центра тяжести кластеров или по «среднему расстоянию» до всех элементов обучающей выборки соответствующих им образов и т.п. [9.7, гл.3].

Следует отметить, что вариабельность изображений объектов в пределах классов такова, что кластеры (локальные сгущения точек в признаковом пространстве), порождаемые конкретными элементами обучающих подмножеств изображений, в определенной степени перекрываются и отношение принадлежности

контролируемого фрагмента изображения (объекта) к тому или иному классу приобретает вероятностный характер. В этих условиях естественно добиваться как можно лучшего качества классификации, что сводится к проблеме синтеза алгоритмов, экстремальных по качеству распознавания. Эта сложная проблема требует комплексного использования теоретических и экспериментальных методов исследования [9.8, 9.9] и в общем случае предполагает знание информативных признаков объектов. Основное требование к таким признакам состоит в том, чтобы они адекватно отражали визуальные (уникальные либо статистические) свойства реальных изображений объектов. Уникальность определяется характерными чертами моделируемых объектов, которые, очевидно, являются признаками их яркостной структуры (*оптической формы*) по которым, собственно, и осуществляется их распознавание.

### 9.3. ВЫЧИСЛЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ

Задача анализа яркостной структуры (формы) объекта чаще всего сводится к аппроксимации ее линейной комбинацией базисных функций. Если функция яркости задана на целочисленной решетке плоскости, то базисные функции получают путем решения алгебраической проблемы собственных значений для матриц изображений объектов. В чем суть этого (и далеко не единственного) подхода?

Пусть  $A - n \times n$  – матрица изображения объекта. Если найдется вектор  $e$ , такой, что верно  $Ae = \lambda e$ , то  $\lambda$  будет называться собственным значением матрицы  $A$ , а  $e$  – соответствующий ему собственный вектор. Если матрица  $A$  нормальная, т. е.  $A^t A = A A^t$ , то она может быть представлена в виде  $A = U \Lambda U^t = \sum_{i=1}^n \lambda_i u_i u_i^t$ , где  $U$  – ортогональная матрица размером

$n \times n$ ,  $u_i$  – столбцы матрицы  $U$ , а  $\Lambda = U^t A U$  – диагональная матрица с собственными значениями матрицы  $A$  на диагонали.

Данное *спектральное разложение* широко используют для распознавания уникальных «матричных» объектов (заданных, например, элементами матрицы  $O$ ), посредством вычисления информативных признаков, которыми служат  $n$  собственных значений  $\Lambda_O = U^t O U$  [9.8, гл.2].

Другой часто используемый на практике тип разложения матриц – *сингулярное разложение* (в английской аббревиатуре SVD –

singular value decomposition). Показано, что любую  $m \times n$  – матрицу  $A$  можно представить в виде  $A = U\Sigma V^t$ , где  $U$  – ортогональная матрица размером  $m \times m$ ,  $V$  – ортогональная  $n \times n$  – матрица, а  $\Sigma$  имеет специальную диагональную форму

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},$$

где  $D = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ ,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  и  $r$  – ранг матрицы  $A$ . Таким образом, SVD матрицы  $A$  можно записать в виде

$$A = U\Sigma V^t = \sum_{i=1}^r \sigma_i u_i v_i^t, \text{ где } u_i - \text{столбцы матрицы } U, \text{ а } v_i - \text{столбцы}$$

матрицы  $V$ . Сингулярные значения  $\sigma_1, \sigma_2, \dots, \sigma_r$  записываются как квадратные корни из ненулевых собственных значений матрицы  $A^t A$ , и их можно использовать для поиска хорошей мало-ранговой аппроксимации исходной матрицы, а следовательно, и в качестве информативных признаков «скрытого образа» объекта.

Коротко изложим также метод построения информативных признаков в статистическом варианте путем разложения по собственным векторам ковариационной матрицы  $\Sigma$  ансамбля наблюдаемых данных. Пусть дано множество  $p \times q$  – матриц изо-

бражений объектов  $\{I_l\}_{l=1}^L$ ; из них мы можем сформировать обучающую выборку векторов-признаков  $\{x_l\}_{l=1}^L$ ,  $x_l \in \mathbb{R}^{p \times q}$ , путем «лексикографического» упорядочения (по столбцам) элементов изображений  $I_l$ . Базисные функции (соответствующие знаменитому разложению Карунена-Лоэва [9.8, гл.8]) получаются путем решения проблемы собственных значений  $\Lambda = \Phi^t \Sigma \Phi$ , где  $\Sigma$  – ковариационная матрица ансамбля исходных данных  $x_l$ ,

$$\Sigma = E(x_l x_l^T) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2m} \\ \cdot & \cdot & \dots & \cdot \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_{mm} \end{pmatrix},$$

где  $\sigma_{ij}$  – ковариация  $i$ -й и  $j$ -й компонент векторов  $x_l$  а  $\sigma_{ii}$  – дисперсия  $i$ -й компоненты вектора измерений  $x_l$ .  $\Phi$  – матрица, составленная из собственных векторов  $\Sigma$ , а  $\Lambda$  – диагональная матрица из собственных значений  $\Sigma$ .

При построении информативного набора признаков исходят из частичного разложения Карунена-Лозва путем идентификации собственных векторов, соответствующих наибольшим собственным значениям (выделение «главных компонент»  $x_i$ ). Это связано с тем, что каждый признак наряду с положительным вкладом в разделение несет в себе в силу ограниченности выборки и шумовую (случайную) составляющую. Если вектор содержит много неинформативных (или малоинформативных) признаков, то отношение «сигнал/шум» (в смысле разделения) значительно лучше для группы высокоинформативных признаков, чем для всей выборки [9.8]. Преобразование  $x_i$  к главным компонентам имеет вид линейного преобразования  $y_i = \Phi_M^T \tilde{x}_i$ , где  $\tilde{x}_i = x_i - \bar{x}$  — центрированный вектор признаков (изображений), а  $\Phi_M$  — подматрица, составленная из главных собственных векторов. Это преобразование выделяет малоразмерное подпространство, соответствующее в базисе Карунена-Лозва максимальным собственным значениям.

#### 9.4. АЛГОРИТМЫ КЛАССИФИКАЦИИ (КРАТКИЙ ОБЗОР)

Варианты основных алгоритмов классификации различают по количеству шагов (этапов) принятия решения, а также по степени и характеру учета статистики признаков. Так, различают одношаговые и многошаговые (последовательные) алгоритмы принятия решений [9.9, гл.4]. В первом варианте принятие решений предусматривает обязательную выдачу оценки  $\hat{i}$  номера класса  $C_i$ , которому принадлежит объект (с приемлемой достоверностью). Если достоверность решения невысока, то целесообразно отказаться от выбора решения на первом шаге, осуществить дополнительный набор признаков и лишь затем принять окончательное решение (или не принять и продолжить процесс наблюдения). Это так называемая последовательная стратегия принятия решения, восходящая к Вальду и его теории последовательного анализа.

По степени учета статистических закономерностей различают *синтаксические* и собственно *статистические* алгоритмы. Из статистических алгоритмов, в свою очередь, выделяют *параметрические* (байесовские и небайесовские), *непараметрические* и *нейрокомпьютерные* алгоритмы. Синтаксические алгоритмы [9.7, гл.8; 9.9, гл.8] вводимые признаки учитывают качественно, часто

двоичными цифрами (0,1) или символами некоторого заданного алфавита. Описание признаков на языке алгебры логики — лингвистическое (кодовое, синтаксическое) служит при этом основой алгоритмов кодирования и распознавания структуры уникальных объектов.

Параметрические байесовские алгоритмы, в отличие от небайесовских учитывают не только статистику распределений значений признаков в классах, но и определенные гипотезы об априорных вероятностях  $P_i$  принадлежности объектов классам. Вероятностную структуру (статистика распределения признаков) объектов распознавания устанавливается путем предварительного натурального эксперимента, математического или физического моделирования. *Введение этой информации можно трактовать как обучение распознаванию, или как адаптацию к конкретным условиям распознавания.*

Непараметрические алгоритмы синтезируют *эвристически* в расчете на неизвестные заранее статистические распределения признаков объектов различных классов. Они используют *локальную* оценку вероятности появления реализации объекта в заданной области по эмпирической частоте (на основе обучающей выборки). Это алгоритмы типа «обобщенной гистограммы» (методы парзеновского окна, ближайших соседей), алгоритмы вычисления оценок, алгебраические алгоритмы и т.д. [9.8, гл.6; 9.11, гл.3–5].

Нейрокомпьютерные алгоритмы отличаются своей заранее заданной универсальной структурой, с большим числом неизвестных параметров, уточняемых в процессе адаптации к входной информации (обучение) [9.10]. Универсальность структуры постулируется из аналогий с гипотетическими нейросистемами (нейросетями). Практические достижения в этой области часто трактуют как проявление возможностей искусственного интеллекта. Обучение (адаптация) нейрокомпьютерных систем (а чаще всего просто специальных алгоритмов) проводится на основе предъявления большого числа обучающих объектов, выбираемых случайно из заданных классов. Статистические закономерности векторов признаков объектов здесь автоматически учитываются в процессе обучения. Признаковое пространство разделяется в результате нелинейными границами на области, соответствующие различным классам. В настоящее время идет рост универсализации структур нейросистем, приводящий к увеличению вычислительных затрат, компенсируемому повышением производительности вычислительных средств.

Далее мы основное внимание уделим лишь параметрическим статистическим методам распознавания. Поэтому перейдем к соответствующему анализу данных в условиях *дискриминантной модели*, известной иначе как модель классификации при наличии обучающих выборок. Здесь используем два вида информации: предположения о свойствах распределений (гладкость, принадлежность к некоторому известному параметрическому семейству) и обучающую выборку.

### 9.5. БАЙЕСОВСКИЙ МЕТОД РАСПОЗНАВАНИЯ

В процессе регистрации объекта и измерения его характерных признаков получают множество чисел, которые составляют вектор наблюдения. Будем считать, что этот вектор наблюдений  $x$  представляет собой случайный вектор с условной плотностью вероятности, зависящей от принадлежности этого вектора определенному классу. При распознавании объектов задачу формально сводят к проверке многих гипотез  $H_1, H_2, \dots, H_k$ , где  $H_i$  — гипотеза, предполагающая принадлежность объекта классу  $C_i$ . Здесь принято, что априорные распределения вероятностей этих гипотез заданы, т. е. известно, с какой вероятностью  $P(H_i)$  объект может принадлежать классу  $C_i$  (или как часто появляется объект данного класса). Причем  $\sum_{i=1}^k P(H_i) = 1$ , поскольку наверняка объект должен принадлежать какому-нибудь классу.

Процесс принятия решений в распознавании объектов можно рассматривать как игру статистического характера, которую классификационный механизм системы распознавания ведет с природой. При каждой реализации игры природа выбирает *стратегию* (в виде состояний природы, соответствующих образам или классам объектов), обозначаемую через  $\{C_i, P(H_i)\}$ . Стратегии игры, применяемые алгоритмом классификации, представляют собой решения, относящиеся к состояниям природы. Каждой паре действий, предпринятой игроками «природа — классификатор» ставится в соответствие некоторая функция потерь (или выигрыша). Считается, что число решений соответствует числу состояний природы (числу классов).

При каждой реализации игры природа выбирает стратегию (класс)  $C_i$  в соответствии с вероятностью  $p(H_i)$ . В результате хода игры, реализованного природой, появляется выборочный

образ (объект)  $x$ . Классификатору не известно, какой именно класс предпочла природа. Вся информация, имеющаяся в его распоряжении, ограничивается самим вектором признаков объекта. Задача классифицирующего механизма — определить, опираясь на эту информацию, к какому классу принадлежит объект  $x$ . Ход игры классификатора, следовательно, представляет собой некоторое решение, определяющее класс  $C_j$ , который («по мнению» классификатора), выбрала природа.

Игры рассматриваемого типа часто называют статистическими. Здесь природа не является «разумным противником», который способен сознательно выбирать свои стратегии таким образом, чтобы добиться максимизации потерь классификатора. Кроме того, у классификатора существует возможность «подсматривать» за игрой природы: он может осуществлять эксперименты и регистрировать обучающее множество объектов, которое затем использует при построении стратегии своей игры.

Пусть при реализации игры между природой и классификатором природа выбирает класс  $C_i$  (стратегию игры) и предъявляет объект  $x$ . Вероятность принадлежности объекта  $x$  классу  $C_i$  обозначим как  $p(H_i | x)$ . Если классификатор принимает решение о том, что объект  $x$  принадлежит классу  $C_j$ , когда на самом деле он принадлежит классу  $C_i$ , то классификатор несет потери, равные  $L_{ij}$ . Так как объект  $x$  может принадлежать любому из  $k$  рассматриваемых классов, то математическое ожидание потерь, связанных с отнесением наблюдаемого объекта к классу  $C_j$ , определяется следующим выражением:

$$\pi_j(x) = \sum_{i=1}^k L_{ij} p(H_i | x); \quad (9.2)$$

в теории статистических решений эту величину часто называют условным средним риском или условными средними потерями.

При распознавании каждого объекта, предъявляемого природой, классификатор может отнести его к одному из  $k$  возможных образов. Если для каждого объекта  $x$  вычисляются значения условных средних потерь  $\pi_1(x), \pi_2(x), \dots, \pi_k(x)$  и классификатор причисляет объект к классу, которому соответствуют наименьшие условные потери, то очевидно, что и математическое ожидание полных потерь на множестве всех решений также будет минимизировано. *Классификатор, минимизирующий математиче-*



ское ожидание общих потерь, называется байесовским [9.8, гл.4]. Со статистической точки зрения байесовский классификатор соответствует оптимальному качеству классификации.

Пусть  $p(\mathbf{x} | H_i)$  есть плотность распределения элементов вектора  $\mathbf{x}$  при условии, что он принадлежит классу  $C_i$ . Хорошо известно, что вероятность принадлежности  $\mathbf{x}$  классу  $C_i$  определяется формулой Байеса

$$p_i = p(H_i | \mathbf{x}) = \frac{P(H_i)p(\mathbf{x} | H_i)}{p(\mathbf{x})}, \quad (9.3)$$

так как безусловная плотность распределения  $p(\mathbf{x}) = \sum_{i=1}^k P(H_i)p(\mathbf{x} | H_i)$ .

Поскольку выражение  $\frac{1}{p(\mathbf{x})}$  входит во все формулы вычисления

условных средних потерь  $\pi_j(\mathbf{x}) = \frac{1}{p(\mathbf{x})} \sum_{i=1}^k L_{ij}P(H_i)p(\mathbf{x} | H_i)$ ,

$j = 1, 2, \dots, k$  в качестве общего множителя, его можно устранить из данного соотношения. В таком случае выражение для средних потерь сводится к следующему:

$$\pi_j(\mathbf{x}) = \sum_{i=1}^k L_{ij}P(H_i)p(\mathbf{x} | H_i), \quad j = 1, 2, \dots, k.$$

При  $k = 2$  и выборе классификатором стратегии (гипотезы)  $H_1$  средние его потери для предъявленного природой объекта  $\mathbf{x}$  равны

$$\pi_1(\mathbf{x}) = L_{11}P(H_1)p(\mathbf{x} | H_1) + L_{21}P(H_2)p(\mathbf{x} | H_2),$$

а при выборе стратегии (гипотезы)  $H_2$  -

$$\pi_2(\mathbf{x}) = L_{12}P(H_1)p(\mathbf{x} | H_1) + L_{22}P(H_2)p(\mathbf{x} | H_2).$$

Как мы знаем, байесовский классификатор обеспечивает отнесение объекта  $\mathbf{x}$  к классу с наименьшим значением средних потерь  $\pi(\mathbf{x})$ . Поэтому объект  $\mathbf{x}$  причисляется к классу  $C_1$ , если выполняется условие  $\pi_1(\mathbf{x}) < \pi_2(\mathbf{x})$ ; это должно означать, что

$$L_{11}P(H_1)p(\mathbf{x} | H_1) + L_{21}P(H_2)p(\mathbf{x} | H_2) < L_{12}P(H_1)p(\mathbf{x} | H_1) + \\ + L_{22}P(H_2)p(\mathbf{x} | H_2),$$

или, что то же самое,

$$(L_{21} - L_{22})p(\mathbf{x} | H_2)p(H_2) < (L_{12} - L_{11})p(\mathbf{x} | H_1)p(H_1). \quad (9.4)$$

Принято считать (и это соответствует здравому смыслу), что потери от ошибочно принятого решения выше «потерь» при правильном выборе. Этому соответствуют неравенства:  $L_{ij} > L_{ji}$ . Тогда байесовское решающее правило (9.4) принимает, естественно, следующий вид:

$$\mathbf{x} \in C_1, \text{ если } \frac{P(H_1)p(\mathbf{x} | H_1)}{P(H_2)p(\mathbf{x} | H_2)} > \frac{L_{21} - L_{22}}{L_{12} - L_{11}}, \text{ или}$$

$$\frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_2)} > \frac{P(H_2)(L_{21} - L_{22})}{P(H_1)(L_{12} - L_{11})}. \quad (9.5)$$

Величину  $\frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_2)}$  называют *отношением правдоподобия* и обо-

значают через  $\Lambda(\mathbf{x}) = \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_2)}$ . Так как  $\Lambda(\mathbf{x})$  представляет собой

отношение двух функций случайной величины, то и само является случайной величиной. Величина в правой части неравенства

(9.5) является пороговым значением  $\eta = \frac{P(H_2)(L_{21} - L_{22})}{P(H_1)(L_{12} - L_{11})}$  *критерия отношения правдоподобия*, к которому в итоге свелось байесовское решающее правило:

$$\mathbf{x} \in C_1, \text{ если } \Lambda(\mathbf{x}) > \eta. \quad (9.6)$$

Отсюда видно, что вся процедура принятия решения сводится к вычислению отношения правдоподобия (зависящего лишь от вектора признаков и параметров распределений классов) и распределение априорных вероятностей или величины потерь на данное отношение  $\Lambda(\mathbf{x})$  влияния не оказывает. Указанная инвариантность процедуры обработки информации имеет большое практическое значение. Часто величины потерь и априорные вероятности являются квалифицированными предположениями на основе предыдущего опыта (интуиции). Неравенство (9.6) позволяет построить решающее правило, рассматривая  $\eta$  как переменный порог, учитывающий изменения в оценках априорных вероятностей и потерь в процессе накопления опыта.

Хорошо известно, что проведение статистического анализа и классификации многомерных наблюдений (признаков природ-

ных объектов) зачастую невозможно ограничить применением некоторых стандартных методов. Необходим детальный анализ структуры наблюдаемой совокупности данных, чтобы путем углубленного исследования представленного числового материала выявить скрытые в нем закономерности, его вероятностную и геометрическую природу. Такой предмодельный (*разведочный*) анализ данных может оказать решающую помощь в компактном и понятном описании структуры наблюдений (например, в форме визуального представления этой структуры — см. рис. 9.2). Отталкиваясь от него, можно «осознанно» поставить вопрос о направлении более детального исследования данных с помощью того или иного метода, а также, возможно, сделать некоторые заключения о причинной модели данных. Если мы хотим использовать двумерное отображение входных данных для понимания внутренней структуры полученной информации с целью последующей классификации, то должны выбрать преобразование (многомерных) данных, сохраняющее разделимость классов. Мы видели, что в случае двух классов отношение правдоподобия несет полную информацию о разделимости классов в байесовском смысле. Следовательно, две плотности вероятностей или монотонные функции от них (отрицательные логарифмы, например) являются подходящей парой переменных. Отображение с использованием этих переменных показано на рис. 9.4. В этом пространстве байесовская граница представляет собой прямую, проходящую под углом  $45^\circ$ , независимо от вида распределений.

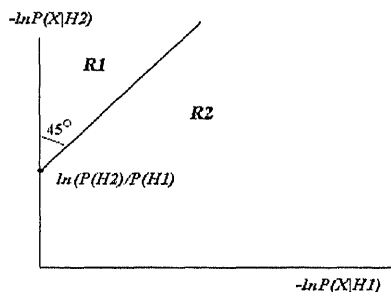


Рис. 9.4. Двумерное отображение данных

Отображение на рис. 9.4 не приводит к потере информации, необходимой для классификации. Единственная сложность здесь — это сложность вычисления функции  $-\ln P(x | H_i)$ . Если плотности вероятностей задаются набором параметров, решение этой

задачи связано с оцениванием параметров. Например, если известно, что плотности вероятностей нормальны, функция  $-\ln P(\mathbf{x} | H_i)$  примет вид

$$-\ln P(\mathbf{x} | H_i) = \frac{1}{2} \left( \mathbf{x} - \hat{\mu}_i \right)^T \hat{\Sigma}_i^{-1} \left( \mathbf{x} - \hat{\mu}_i \right) + \frac{1}{2} \ln \left\{ (2\pi)^m \left| \hat{\Sigma}_i \right| \right\},$$

где  $\hat{\mu}_i$  и  $\hat{\Sigma}_i$  — оценки математического ожидания и ковариационной матрицы. В непараметрическом случае можно воспользоваться одним из множества методов оценивания плотностей (разложением по базисным функциям, например [9.7, гл.7]).

**Пример 9.1.** Пусть при гипотезе  $H_1$  наблюдаемый фрагмент изображения соответствует постоянному «фону» с яркостью  $b_1 > 0$ , а по гипотезе  $H_2$  фрагмент изображения соответствует «объекту» с постоянной яркостью  $b_2 > b_1$  (рис. 9.5). Наблюдаемый яркостный сигнал подвержен шумовым искажениям. Будем считать, что фрагмент однородный по яркости и содержит  $N$  отсчетов. Результаты наблюдений представляют ряд из  $N$  независимых гауссовых величин  $x_1, x_2, \dots, x_N$  с известным средним значением: либо  $b_1$  при гипотезе  $H_1$ , либо  $b_2$  при гипотезе  $H_2$ . Вследствие статистической независимости нетрудно записать совместные плотности вероятности величин  $x_j, j = 1, 2, \dots, N$  при каждой из гипотез:

$$P(\mathbf{x} | H_i) = \prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x_j - b_i)^2}{2\sigma^2} \right),$$

где  $\sigma^2$  — известная дисперсия шума. Критерий отношения правдоподобия в этом случае имеет простой вид:

$$\Lambda(\mathbf{x}) = \frac{\prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x_j - b_1)^2}{2\sigma^2} \right)}{\prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(x_j - b_2)^2}{2\sigma^2} \right)}.$$

После приведения подобных членов и взятия логарифма получим

$$\ln \Lambda(\mathbf{x}) = \frac{b_1 - b_2}{\sigma^2} \sum_{j=1}^N x_j + \frac{N(b_2^2 - b_1^2)}{2\sigma^2},$$

и критерий Байеса запишется в виде  $x \in C_1$ , если

$$\frac{b_1 - b_2}{\sigma^2} \sum_{j=1}^N x_j + \frac{N(b_2^2 - b_1^2)}{2\sigma^2} > \ln \eta,$$

или в эквивалентной форме

$$\sum_{j=1}^N x_j < \frac{\sigma^2}{b_2 - b_1} \ln \eta + \frac{N(b_2 + b_1)}{2}.$$



Рис. 9.5. Изображение сцены в инфракрасном диапазоне (источник [9.16]). Рамками отмечены фоновый фрагмент (вверху) и фрагмент, соответствующий изображению объекта (яркая область повышенной теплоотдачи работающего двигателя)

Нетрудно видеть, что процедура классификации сводится просто к суммированию результатов наблюдений яркости на распознаваемом фрагменте и сравнению суммы с порогом

$$\gamma = \frac{\sigma^2}{b_2 - b_1} \ln \eta + \frac{N(b_1 + b_2)}{2}.$$

Так как реализация байесовского классификатора предполагает знание плотности распределения для каждого класса, то становится совершенно очевидным, что оценка плотностей — основная проблема такой схемы классификации. С точки зрения

статистического анализа при выборе модели распределений вполне обоснованным представляется принцип: среди множества моделей следует использовать модель, которая позволяет делать максимально надежные выводы о лежащей в основе структуры данных статистике (функции данных). Общим выражением этого положения является *принцип максимальной энтропии* (МЭ), который гласит: если мы делаем выводы по неполной информации, то должны опираться на такое распределение вероятностей, которое имеет максимальную энтропию, допускаемую нашей априорной информацией. При этом признается, что модельное распределение с более высокой энтропией в некотором смысле «предпочтительнее» модели распределения с малой энтропией (позволяет извлечь больше информации из наблюдаемых данных). Поэтому можно показать, что если исходная информация о параметрах модели распределения состоит лишь из вектора средних значений (признаков) и элементов ковариационной матрицы, то распределением с максимальной энтропией является многомерное нормальное [9.12]. Принцип МЭ строит Гауссовскую форму как распределение, которое может быть реализовано «Природой» наибольшим числом способов (*можно считать это оптимальной стратегией природы в статистической игре*), совместимых с исходными данными (средними и ковариациями выборки).

**Пример 9.2.** Рассмотрим модель двух многомерных нормальных совокупностей с равными ковариационными матрицами  $\Sigma$ ,  $\mathcal{N}(\bar{\mu}_1, \Sigma)$  и  $\mathcal{N}(\bar{\mu}_2, \Sigma)$ , где  $\bar{\mu}_1$  и  $\bar{\mu}_2$  — вектора средних значений классов. Соответствующие (предполагающиеся заданными) ковариационные матрицы имеют вид

$$\Sigma = E(\mathbf{x}\mathbf{x}^t) = \{\sigma_{ij}\}_{i,j=1}^m,$$

где  $\sigma_{ij}$  — ковариация  $i$ -й и  $j$ -й компонент вектора признаков  $\mathbf{x}$ , а  $\sigma_{ii}$  — дисперсия  $i$ -й компоненты вектора измерений  $\mathbf{x}$ . Поскольку в случае нормального распределения имеем

$$p(\mathbf{x} | H_i) = (2\pi)^{-m/2} \left( \Sigma^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{x} - \bar{\mu}_i)^T \Sigma^{-1} (\mathbf{x} - \bar{\mu}_i) \right] \right), \quad (9.7)$$

( $\bar{\mu}_i$  — известный вектор математического ожидания  $\mathbf{x}$  при гипотезе  $H_i$ ), то отношение двух плотностей для байесовского правила (9.5) определяется выражением

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_2)} = \exp \left\{ -\frac{1}{2} \left[ (\mathbf{x} - \bar{\mu}_1)^T \Sigma^{-1} (\mathbf{x} - \bar{\mu}_1) - (\mathbf{x} - \bar{\mu}_2)^T \Sigma^{-1} (\mathbf{x} - \bar{\mu}_2) \right] \right\}. \quad (9.8)$$

Областью  $R_1$ , при попадании в которую наблюдение классифицируется как принадлежащее первому классу, является множество векторов  $\mathbf{x}$ , для которых величина  $\Lambda(\mathbf{x})$  больше некоторой константы  $\eta$ , выбираемой подходящим способом. В частности, при выборе двоичной функции потерь ( $L_{11} = L_{22} = 0$  — правильное решения,  $L_{12} = L_{21} = 1$  — ошибка) условие, определяющее принадлежность образа  $\mathbf{x}$  классу  $C_1$  принимает вид

$\ln \Lambda(\mathbf{x}) = \ln \left( \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_2)} \right) > \ln \eta$ , где  $\ln \eta = \ln \left( \frac{P(H_2)}{P(H_1)} \right)$ , если  $\mathbf{x} \in C_1$  и известны априорные вероятности  $\{P(H_1), P(H_2)\}$ . Группируя соответствующие члены представления  $\ln \Lambda$ , получаем

$$\ln \Lambda = \mathbf{x}^t \Sigma^{-1} (\bar{\mu}_1 - \bar{\mu}_2) - \frac{1}{2} (\bar{\mu}_1 + \bar{\mu}_2)^t \Sigma^{-1} (\bar{\mu}_1 - \bar{\mu}_2) = U. \quad (9.9)$$

Первый член данного выражения носит название собственно дискриминантной (решающей) функции  $d(\mathbf{x}) = \mathbf{x}^t \Sigma^{-1} (\bar{\mu}_1 - \bar{\mu}_2)$ . В результате область  $R_1$  определяется линейной разделяющей функцией следующим образом:

$$R_1 : d(\mathbf{x}) \equiv \mathbf{x}^t \Sigma^{-1} (\bar{\mu}_1 - \bar{\mu}_2) > \frac{1}{2} (\bar{\mu}_1 + \bar{\mu}_2)^t \Sigma^{-1} (\bar{\mu}_1 - \bar{\mu}_2) - \ln \left( \frac{P(H_1)}{P(H_2)} \right) = \eta_1. \quad (9.10)$$

Рассмотрим случай, когда два многомерных нормальных распределения имеют разные ковариационные матрицы. Решающее правило приобретает вид

$$\mathbf{x} \in C_1, \text{ если } : d(\mathbf{x}) = (\mathbf{x} - \mu_1^T) \Sigma_1^{-1} (\mathbf{x} - \mu_1) - (\mathbf{x} - \mu_2^T) \Sigma_2^{-1} (\mathbf{x} - \mu_2) + \ln \frac{|\Sigma_1|}{|\Sigma_2|} \leq 2 \ln \frac{P(H_1)}{P(H_2)}.$$

Видно, что решающая граница является квадратичной формой относительно вектора признаков (рис. 9.6, а). Для таких ситуаций рекомендован простой подход к синтезу линейной разделяющей функции [9.8]. Он состоит в замене каждой из ковариационных матриц их средним значением, т. е.  $P(H_1)\Sigma_1 + P(H_2)\Sigma_2$ .

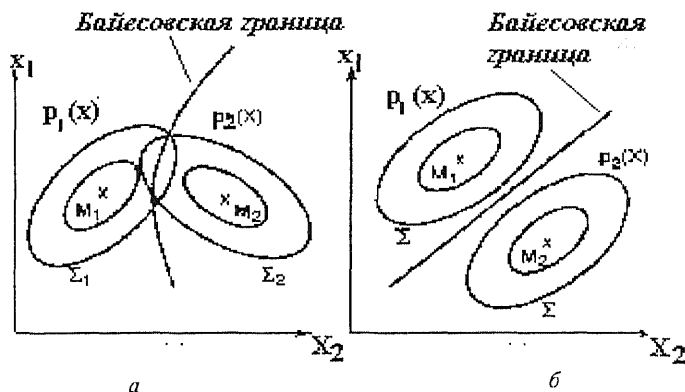


Рис. 9.6. Решающие границы для нормальных распределений:

а -  $\Sigma_1 \neq \Sigma_2$ ; б -  $\Sigma_1 = \Sigma_2$

**Пример 9.3.** Проанализируем сейчас практически более важную ситуацию, когда имеются две обучающие выборки —  $\{(\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_{n_1}^1), (\mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_{n_2}^2)\}$ , но параметры соответствующих им классов (нормальных распределений) неизвестны. На основе этой информации необходимо классифицировать вектор признаков  $\mathbf{x}$  неизвестного объекта. Очевидно, наилучшими оценками  $\hat{\mu}_j, j = 1, 2$  являются статистики  $\hat{\mu}_j = \bar{\mathbf{x}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{x}_i^j$ , а эффективной оценкой ковариационной матрицы  $\Sigma$  является матрица  $S$ , определяемая в виде

$$S = \hat{\Sigma} = \frac{1}{n_1 + n_2 - 2} \left[ \sum_{j=1}^2 \sum_{i=1}^{n_j} (\mathbf{x}_i^j - \bar{\mathbf{x}}_j)(\mathbf{x}_i^j - \bar{\mathbf{x}}_j)^T \right].$$

Тогда оценка дискриминантной функции, полученная по обучающим выборкам, имеет вид

$$d(\mathbf{x}) = \mathbf{x}^t S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2).$$



Это линейная функция (случайная величина), имеющая наибольшую дисперсию между выборками относительно дисперсии внутри выборок. Чаще всего в качестве статистики критерия классификации привлекают следующую величину (так называемая *модель Фишера*) [9.13]

$$\begin{aligned}\hat{U}(\mathbf{x}) &= \mathbf{x}' \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = \\ &= \left[ \mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right]' \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2).\end{aligned}$$

### 9.6. ВЕРОЯТНОСТИ ОШИБОК КЛАССИФИКАЦИИ

Из примера 2 ясно, что вероятность неправильной классификации образа, принадлежащего классу  $C_1$ , равна  $P(U < \ln \eta \mid H_1)$ , а вероятность неправильной классификации образа, принадлежащего классу  $C_2$ , равна  $P(U > \ln \eta \mid H_2)$ . Если априорные вероятности неизвестны, то  $\eta_1 = \ln \eta$  можно выбрать из условия, чтобы математическое ожидание потерь, связанных с ошибочным отношением среди пары классов были равны. Для этого нужно найти распределение случайной величины

$$U = \mathbf{x}' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2) - \frac{1}{2}(\bar{\boldsymbol{\mu}}_1 + \bar{\boldsymbol{\mu}}_2)' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)$$

для каждого из двух возможных состояний природы. Если вектор признаков принадлежит первому классу (т. е.  $E(\mathbf{x}) = \bar{\boldsymbol{\mu}}_1$ ), то величина  $U$  распределена нормально с математическим ожиданием

$$\begin{aligned}E(U) &= \bar{\boldsymbol{\mu}}_1' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2) - \frac{1}{2}(\bar{\boldsymbol{\mu}}_1 + \bar{\boldsymbol{\mu}}_2)' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2) = \\ &= \frac{1}{2}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2) = \frac{\Delta}{2},\end{aligned}$$

где  $\Delta = (\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)$  имеет смысл «нормированного расстояния» между центрами классов и называется расстоянием *Махаланобиса*. Дисперсия  $U$  также равна  $\Delta$ :

$$\sigma^2(U) = E\left\{(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \bar{\boldsymbol{\mu}}_1)(\mathbf{x} - \bar{\boldsymbol{\mu}}_1)' \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{\mu}}_1 - \bar{\boldsymbol{\mu}}_2)\right\} = \Delta.$$

Таким образом, если  $x \sim \mathcal{N}(\bar{\mu}_1, \Sigma)$  то  $U$  распределена по нормальному закону  $\mathcal{N}(\frac{\Delta}{2}, \Delta)$ . Легко видеть, что если  $x \sim \mathcal{N}(\bar{\mu}_2, \Sigma)$ , то  $U$  распределена по нормальному закону  $\mathcal{N}(-\frac{\Delta}{2}, \Delta)$ . Вероятность ошибочной классификации при условии, что  $x \sim \mathcal{N}(\bar{\mu}_1, \Sigma)$ , равна

$$\begin{aligned} P(C_2 | H_1) &= \int_{-\infty}^{\eta_1} \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{1}{2}(u-\frac{\Delta}{2})^2 / \Delta} du = \\ &= \int_{-\infty}^{(\eta_1 - \frac{\Delta}{2}) / \sqrt{\Delta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \Phi(\eta_1 - \frac{\Delta}{2}), \end{aligned} \quad (9.11)$$

где  $\Phi(u)$  — функция стандартного нормального распределения  $\mathcal{N}(0, 1)$ . Аналогично, вероятность ошибочной классификации при условии, что наблюдение  $x \sim \mathcal{N}(\bar{\mu}_2, \Sigma)$ , равна (см. рис. 9.7):

$$P(C_1 | H_2) = \int_{\eta_1}^{\infty} \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{1}{2}(u+\frac{\Delta}{2})^2 / \Delta} du = \int_{(\eta_1 + \frac{\Delta}{2}) / \sqrt{\Delta}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy. \quad (9.12)$$

Если  $\eta_1 = 0$ , то  $P(C_1 | H_2) = P(C_2 | H_1)$ , и общая вероятность ошибочной классификации равна

$$P_{\text{error}} = \int_{\frac{\Delta}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy. \quad (9.13)$$

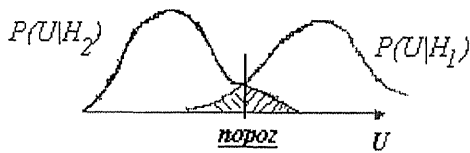


Рис. 9.7. Вероятность ошибочной классификации (заштрихованные области)

Если оценки параметров классов неизвестны (см. модель Фишера), то вероятность ошибочной классификации здесь будет случайной величиной, зависящей от выборок, их объемов ( $n_1, n_2$ ) и

размерности признакового пространства  $m$ . Для изучения данной задачи А.Н. Колмогоровым была предложена асимптотика, в которой  $(n_j, m) \rightarrow \infty$  так, что  $\frac{m}{n_j} = \lambda_j = \text{const}$ .

В этой постановке модель Фишера была изучена Деевым [9.14] в двух вариантах — при *известной* ковариационной матрице  $\Sigma$ , и *оцениваемой* по паре обучающих выборок. В частности, в первом случае асимптотическое значение ошибки классификации

$$P_{\text{error}}^1 \Rightarrow \Phi \left( -\frac{\Delta}{2\sqrt{\Delta + \lambda_1 + \lambda_2}} \right),$$

где  $\Delta$  есть предел (при  $m \rightarrow \infty$ ) расстояния Махаланобиса  $\Delta_m = (\bar{\mu}_1 - \bar{\mu}_2)' \Sigma^{-1} (\bar{\mu}_1 - \bar{\mu}_2)$  между классами. Во втором варианте

$$P_{\text{error}}^2 \Rightarrow \Phi \left( -\frac{\Delta \left( 1 - \lambda_1 \lambda_2 / (\lambda_1 + \lambda_2) \right)^{\frac{1}{2}}}{2\sqrt{\Delta + \lambda_1 + \lambda_2}} \right).$$

Из сравнения двух формул для ошибок классификации видно, что цена, которую приходится платить за  $m(m+1)/2$  неизвестных элементов ковариационной матрицы, достаточно высока. В связи с этим правильно проводить классификацию исходя из предположения равенства ковариационных матриц  $\Sigma_1 = \Sigma_2 = \Sigma$  (иначе добавляется еще  $m(m+1)/2$  неизвестных параметров). Кроме того, рекомендуется при оценивании  $P_{\text{error}}^2$  подставлять вместо параметра  $\Delta$  его несмещенную оценку

$$\bar{\Delta} = \frac{n-m-1}{n} \hat{\Delta} - m \frac{n_1 + n_2}{n_1 n_2},$$

где  $\hat{\Delta} = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$ ,  $n = n_1 + n_2 - 2$ .

Анализ данной формулы показывает, что при постоянном расстоянии Махаланобиса  $\Delta$  средняя ошибка увеличивается с ростом числа признаков  $m$ . Следовательно, решающие правила в задачах классификации не должны использовать избыточных признаков, так как каждый лишний (малоинформативный) признак может существенно увеличить ошибку.

Оценка ошибки по обучающей выборке (в асимптотике Колмогорова-Деева) дает, в общем, чересчур оптимистичный результат; доли ошибок чаще всего получаются довольно незначительными. Самый простой путь для получения несмещенной оценки ошибки — разделить имеющиеся выборки из отдельных групп на две части. Затем одну часть рассматривать как *обучающую* выборку, используемую лишь для построения решающего правила (оценивания  $\hat{\mu}_j, \Sigma$ ), а другую — как *контрольную*, т. е. только для оценивания ошибок классификации. Но поскольку объемы выборок при этом сильно уменьшаются, этот метод дает большую ошибку в оценке искомых вероятностей. Лаченбрух [9.8, гл.5] предложил компромиссный вариант — *метод исключения одного объекта* (известный в отечественной литературе как метод «скользящего контроля»). Метод состоит из  $n = n_1 + n_2 + \dots + n_k$  шагов, в общем соответствующих объему обучающей выборки из  $n$  векторов наблюдений всех  $k$  классов. При каждом шаге дискриминантное правило строится по множеству из  $n - 1$  векторов наблюдений, а затем проверяется на оставшемся (в качестве контрольного) исключенном наблюдении. Доля  $v$  неправильно классифицированных при этом векторов признаков в  $n$  шагах скользящего контроля дает (приблизительно) вероятность ошибок решающего правила  $v/n$ . Данный метод приводит в целом к несколько завышенной оценке вероятности ошибки классификации. В то же время если для проектирования классификатора использовать полностью обучающие множества ( $n$  векторов признаков), то проверка его качества по выбранному методу дает заниженную оценку вероятности ошибки. Установлено [9.8], что смещение между данными оценками ВО уменьшается приблизительно пропорционально  $1/n$ , а среднеквадратичное отклонение оценок ВО уменьшается как  $(1/n)^2$ .

### 9.7. ПРОБЛЕМА К-КЛАССОВ

Перейдем к задаче классификации объекта  $x$  при задании  $k$ -классов, формулируя критерий отношения правдоподобия в виде хорошо известного в математической статистике «критерия значимости». С его помощью проверяем одновременно и принадлежность вектора  $x$  к классу  $C_j$ . При построении критерия вектор  $x$  в этом варианте рассматривается как представитель класса с номером 0 (т. е.  $C_0$ ) наряду с классами  $C_1, C_2, \dots, C_k$ .

Тогда в соответствии с изложенным выше, статистика дискриминантного критерия представляется в виде случайной переменной из  $F$  – распределения [9.13, п.7-7]

$$d_{0j} = d_j = \left( \frac{n-k-m+1}{m(n-k)} \right) \frac{n_j}{n_j+1} (\mathbf{x} - \bar{\mathbf{x}}_j)^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_j).$$

Напомним, что  $F$ -распределение с  $(v_1, v_2)$  степенями свободы имеет плотность

$$F_{v_1, v_2}(y) = \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{v_1/2}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right)} y^{\frac{v_1}{2}-1} \left(1 + \frac{v_1}{v_2} y\right)^{-\frac{v_1+v_2}{2}}.$$

Вектор  $\mathbf{x}$  (т. е. соответствующий ему объект) здесь относится к классу  $C_j$  и принимается гипотеза  $H_{0j} : \bar{\mu}_0 = \bar{\mu}_j$ , если  $d_j \leq F_{m, n-m-k+1; \alpha}$ , где  $F_{m, n-m-k+1; \alpha}$  – квантиль  $F$ -распределения с  $(m, n-m+1)$  степенями свободы. Гипотеза отвергается, (т. е.  $\bar{\mu}_0 - \bar{\mu}_j \neq 0$ ), если  $d_j > F_{m, n-m-k+1; \alpha}$ . Здесь  $F_{v_1, v_2; \alpha}$  – критическое (пороговое) значение, которое можно найти в таблицах квантилей  $F$ -распределения для каждого заданного уровня значимости  $\alpha$  (характеризующего вероятность непринятия гипотезы  $H_{0j} : \bar{\mu}_0 = \bar{\mu}_j$ , когда она верна). Следует отметить, что при использовании данного метода классификации объект может быть отнесен одновременно к нескольким классам. Но может случиться, что он не будет причислен ни к одному классу. В первом случае можно выбрать класс с наименьшим значением  $d_j$  (наиболее вероятное решение). Благодаря неравенству  $d_j \leq F_{m, n-m-k+1; \alpha}$  каждому классу соответствует  $m$ -мерный шар, с вероятностью  $1-\alpha$  содержащий объекты, действительно относящиеся к этому классу. В частности, при  $\alpha = 0,05$  95 % объектов располагаются внутри области рассеяния своего класса. Классификацию можно провести точнее, если дополнительно учитывать априорные вероятности  $P(H_j)$  попадания объекта в каждый класс.

**КОНТРОЛЬНЫЕ ВОПРОСЫ К ГЛАВЕ 9**

9.1. Дайте содержательную формулировку проблемы распознавания объектов. В чем разница между процедурами распознавания и классификации?

9.2. Какую роль выполняют разделяющие функции в распознавании объектов?

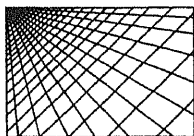
9.3. Охарактеризуйте алгебраические методы вычисления информативных признаков матричных объектов.

9.4. Выделите основные подходы к классификации в зависимости от наличия обучающего материала и информации о структуре исходных данных.

9.5. Выведите байесовское решающее правило.

9.6. Постройте байесовский классификатор для нормальных распределений при известных и неизвестных параметрах классов.

9.7. Опишите методы оценивания вероятностей ошибок при статистической классификации.



## ГЛАВА 10

### МАТЕМАТИЧЕСКАЯ МОРФОЛОГИЯ И ОБРАБОТКА ИЗОБРАЖЕНИЙ

Одним из сравнительно новых направлений в анализе изображений является применение аппарата математической морфологии. Начало математической морфологии, использующей представления теории множеств и интегральной геометрии, было положено работами французских исследователей Ж. Матерона [10.1] и Дж. Серра [10.2], занимавшихся проблемами минералогии и петрографии. Цель их исследований состояла в количественном описании физических и механических свойств материалов посредством анализа их геометрической структуры. Затем математическая морфология достигла состояния серьезного инструмента обработки изображений с основным применением в материаловедении, исследовании цитологических препаратов, анализе медицинских изображений.

Конечно, объема одной главы совершенно недостаточно для сколь-нибудь последовательного изложения теоретических основ, поэтому она имеет скорее иллюстративный характер. Здесь фрагментарно обсуждаются основные операции математической морфологии и их свойства и приводятся результаты применения этих операций для обработки и анализа изображений (в основном двухградационных).

Следует заметить, что публикации, посвященные как теоретическим вопросам математической морфологии, так и ее приложениям в области обработки изображений, в русскоязычной литературе практически отсутствуют. При написании этого материала у нас возникали трудности с некоторыми русскоязычными названиями морфологических операций, адекватно передающими смысл названий, введенных в оригинальных англоязычных работах [10.2, 10.3], на которых базируется изложение. Обозначения в основном совпадают с принятыми в [10.2, 10.3].

Напомним некоторые основные понятия из теории множеств, которые потребуются в дальнейшем. Пусть  $E^n$  —  $n$ -мерное пространство. Ниже примем, что  $E^n = \mathbf{R}^n$  или  $E^n = \mathbf{Z}^n$ , где  $\mathbf{R}^n$  —  $n$ -мерное евклидово пространство, а  $\mathbf{Z}^n$  —  $n$ -мерное дискретное пространство ( $n$ -мерная решетка). В применении к изображениям, как правило, рассматриваются двумерные пространства. Если  $X = \{x\}$  и  $Y = \{y\}$  — множества в  $E^n$ , то *объединением множеств*  $X$  и  $Y$  называется множество  $Z = \{z : z \in X \text{ или } z \in Y\} = X \cup Y$ , (т.е. множество, состоящее из таких элементов  $z$ , которые принадлежат  $X$  или  $Y$ ), а *пересечением множеств*  $X$  и  $Y$  называется множество  $Z = X \cap Y = \{z : z \in X, z \in Y\}$ . Множество  $Z = X^c = \{z : z \notin X\}$  называется *дополнением множества*  $X$ . *Разностью множеств*  $X$  и  $Y$  называется множество  $Z = \{z : z \in X, z \notin Y\} = X \setminus Y$ . Множество  $Z$  называется пустым, если оно не содержит ни одного элемента. Обозначается такое множество как  $Z = \emptyset$ . Справедливы следующие соотношения:

$$\begin{aligned}(X \cup Y)^c &= X^c \cap Y^c; \\ (X \cap Y)^c &= X^c \cup Y^c; \\ X \setminus Y &= X \cap Y^c.\end{aligned}\tag{10.1}$$

Определим на  $E^n$  *индикаторную функцию* множества  $X$  следующим образом:

$$\varphi_X(x) = \begin{cases} 1, & x \in X \\ 0, & x \notin X \end{cases}.$$

Определим также меру множества  $X$ :

$$\text{Mes}(X) = \int_{\mathbf{R}^n} \varphi_X(x) dx \quad \text{— для непрерывного пространства } \mathbf{R}^n \text{ и}$$

$$\text{Mes}(X) = \sum_{x_i \in X} \varphi_X(x_i) \quad \text{— для дискретного пространства } \mathbf{Z}^n.$$

Для изображений эти определения означают, что мерой множества  $X$  является его площадь в непрерывном случае и количество узлов решетки, входящих в множество, — в дискретном.



## 10.1. ОПЕРАЦИИ МАТЕМАТИЧЕСКОЙ МОРФОЛОГИИ

Двухградационное изображение можно рассматривать как индикаторную функцию набора множеств в  $E^2$  (как, например, индикаторную функцию множества  $X_1 \cup X_2 \cup X_3$  на рис. 10.1). Для данного множества  $X$  можно зафиксировать некоторый элемент (не обязательно принадлежащий этому множеству), который назовем центром (или началом) множества  $X$ . Обозначим через  $X_a$  множество  $X$ , центр которого помещен в точку  $a$ . Одним из основных понятий математической морфологии является понятие *структурного элемента*. Структурный элемент  $B$  – это множество, состоящее из двух пересекающихся подмножеств  $B^1$  и  $B^2$ , для которых определено общее начало.

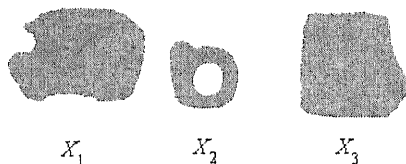


Рис. 10.1. Двухградационное изображение

## НМ-преобразование

Согласно [10.2] базовым преобразованием, позволяющим строить набор различных операций математической морфологии, является преобразование Hit or Miss. Нам не удалось найти адекватного перевода этому названию, поэтому далее будем пользоваться названием НМ-преобразование. Для данного множества  $X = \{x \in E^n\}$  и данного структурного элемента  $B$  результат НМ-преобразования определяется как

$$Y = X \otimes B = \{x : B_x^1 \subset X, B_x^2 \subset X^c\}. \quad (10.2)$$

(Здесь через  $X^c$  обозначено дополнение множества  $X$ .)

Нетрудно видеть (рис. 10.2), что в результате НМ-преобразования на исходном изображении выделяются элементы, окрестность которых совпадает со структурным элементом (заметим, что форма окрестности определяется формой структурного элемента). Условие (10.2) выполняется для элементов, лежащих на нижней границе  $X$  (например, 1...4 позиции структурного элемента). В позиции 5  $B_x^1 \subset X$ , но  $B_x^2 \not\subset X^c$ , в позиции 6, наоборот,  $B_x^1 \not\subset X$ , но  $B_x^2 \subset X^c$ , а в позиции 7 не выполняются оба условия.

Применяя НМ-преобразование с различными структурными элементами, можно выделять специфические геометрические особенности изображений.

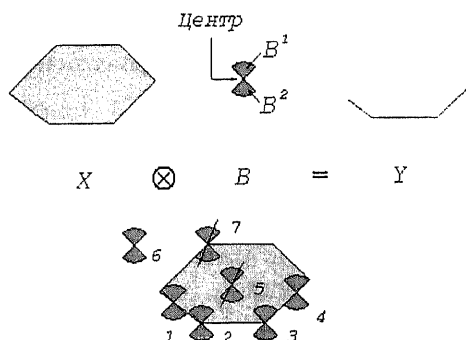


Рис. 10.2. НМ-преобразование

### Эрозия

Частным случаем НМ-преобразования является операция *эрозии* (*erosion*). Пусть в структурном элементе  $B$  подмножество  $B^2$  — пусто ( $B^2 = \emptyset$ ). При этом условии  $B_x^2 \subset X^c$  всегда выполняется и в множество  $Y$  включаются только те элементы исходного множества  $X$ , для которых выполняется условие  $B_x \subset X$ :

$$Y = X \ominus B = \{x : B_x \subset X\}. \quad (10.3)$$

Иначе говоря, если  $x \in X$ , а  $b \in B$ , то в множество  $Y$  включаются такие элементы, для которых выполняется условие  $x+b \in X$  (рис. 10.3).

С другой стороны, если  $b$  пробегает все возможные положения в  $B$ , условие  $x+b \in X$  выполняется тогда и только тогда, когда  $x$  принадлежит смещенному множеству  $X_{-b}$  (рис. 10.4). Поэтому другое, эквивалентное, представление операции эрозии имеет вид

$$Y = X \ominus B = \bigcap_{b \in B} X_{-b} = \bigcap_{-b \in B} X_b = \bigcap_{b \in \bar{B}} X_b, \quad (10.3')$$

где  $\bar{B} = \{b : -b \in B\}$  — множество, симметричное  $B$  относительно его начала. Это представление может оказаться полезным при численной реализации операции эрозии.

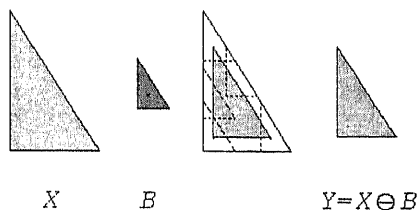


Рис. 10.3. Эрозия

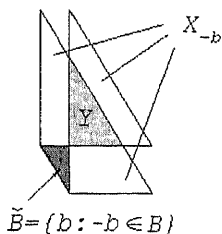


Рис. 10.4. Эрозия как пересечение смещенных множеств

### Дилатация

Операцией, двойственной к эрозии, является *дилатация* (*dilation*), которая определяется следующим образом (рис. 10.5):

$$Y = X \oplus B = \{x : B_x \cap X \neq \emptyset\}. \quad (10.4)$$

Другое представление дилатации имеет вид

$$Y = X \oplus B = \bigcup_{b \in \tilde{B}} X_b, \quad (10.4')$$

как это показано на рис. 10.6.

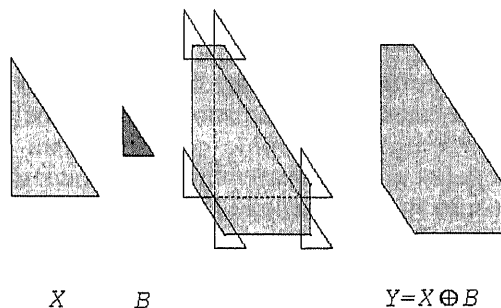


Рис. 10.5. Дилатация

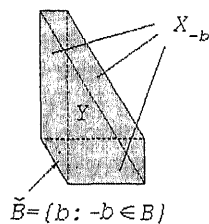


Рис. 10.6. Дилатация как объединение смещенных множеств

Если рассматривать множество  $X$  как объект, а  $X^c$  как фон в изображении, то дилатацию объекта можно интерпретировать как эрозию фона:

$$X \oplus B = (X^c \ominus B)^c. \quad (10.5)$$

Действительно,

$$(X^c \ominus B)^c = \left( \bigcap_{b \in B} X_b^c \right)^c = \bigcup_{b \in B} X_b = X \oplus B.$$

### Алгебраические свойства дилатации и эрозии

Приведем здесь без доказательства ряд полезных свойств рассмотренных операций.

а) Дистрибутивность: дилатация дистрибутивна относительно объединения

$$\left( \bigcup_{i=1}^I X_i \right) \oplus B = \bigcup_{i=1}^I (X_i \oplus B), \quad (10.6)$$

а эрозия — относительно пересечения множеств

$$\left( \bigcap_{i=1}^I X_i \ominus B \right) = \bigcap_{i=1}^I (X_i \ominus B). \quad (10.6')$$

Свойство дистрибутивности с учетом соотношения (10.5) позволяет выполнять операции над  $X$  по фрагментам, комбинируя затем результаты посредством объединения или пересечения.

б) Итеративность:

$$(X \oplus A) \oplus B = X \oplus (A \oplus B); \quad (10.7)$$

$$(X \ominus A) \ominus B = X \ominus (A \oplus B). \quad (10.7')$$

Это чрезвычайно важное свойство, поскольку оно позволяет разлагать сложные структурные элементы в композицию более простых (рис. 10.7). Соответственно, операции со сложными элементами могут быть заменены последовательностью операций с более простыми. Так, эрозию некоего множества  $X$  посредством структурного элемента  $H$ , приведенного на рис. 10.7, можно заменить четырьмя последовательными эрозиями со структурными элементами  $A - D$ .

в) инвариантность к изменению масштаба:

$$\lambda X \oplus \lambda B = \lambda (X \oplus B); \quad (10.8)$$

$$\lambda X \ominus \lambda B = \lambda (X \ominus B). \quad (10.8')$$

В этих соотношениях через  $\lambda B$ ,  $\lambda X$  обозначены множества

$$B' = \lambda B = \{b' = \lambda b : b \in B\} \text{ и } X' = \lambda X = \{x' = \lambda x : x \in X\}$$

(рис. 10.8).

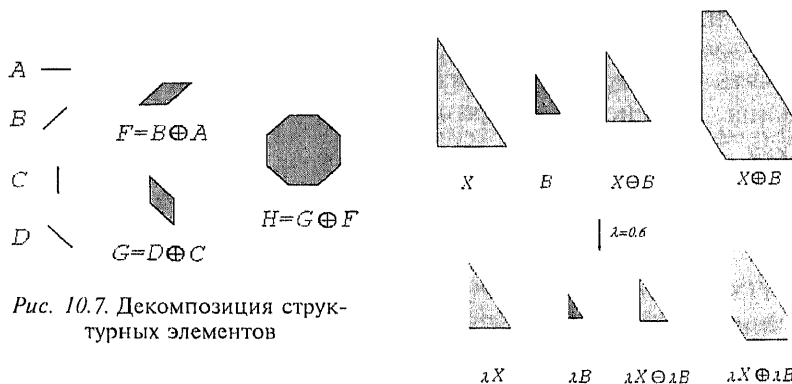


Рис. 10.7. Декомпозиция структурных элементов

Рис. 10.8. Инвариантность эрозии и дилатации к масштабным преобразованиям

### Применение эрозии и дилатации

Эрозия и дилатация — операции, предназначенные в первую очередь для выявления различных морфологических особенностей изображений, с использованием различных структурных элементов. Например, эрозия посредством круга с радиусом  $r$  позволяет найти в изображении объекты, минимальный поперечный размер которых превышает  $2r$ . Если же в качестве структурного элемента взять две точки, смещение между которыми определяется вектором  $\mathbf{h}$ , эрозия позволит выделить объекты, имеющие соседей в направлении и на расстоянии, заданных этим вектором (рис. 10.9). (Под объектами здесь подразумеваются односвязные множества.)

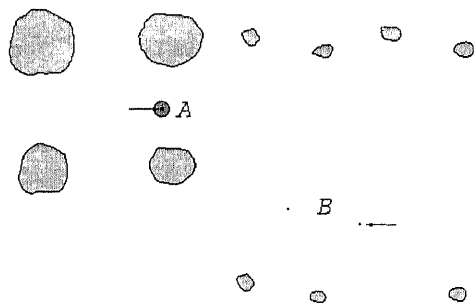


Рис. 10.9. Верхний ряд — исходное множество  $X$ . Средний ряд —  $X \ominus A$ . Нижний ряд —  $X \oplus B$ . Стрелками помечены начала структурных элементов

Более интересное применение эрозии с двухточечным структурным элементом заключается в том, что с ее помощью можно вычислять автокорреляцию изображения. Автокорреляция изображения, заданного индикаторной функцией  $\varphi_X(x)$ , определяется как

$$K_X(h) = \int_{\mathbb{R}^n} \varphi_X(x) \varphi_X(x+h) dx = \int_{\mathbb{R}^n} \psi_Y(x; h) dx,$$

где  $\psi_Y(x; h)$  можно интерпретировать как индикаторную функцию множества  $Y = X \cap X_{-h}$ , зависящего от параметра  $h$ , поскольку

$$\psi_Y(x; h) = \begin{cases} 1, & x \in Y = X \cap X_{-h} \\ 0, & x \notin Y \end{cases}.$$

Нетрудно убедиться, что  $X \ominus B = X \cap X_{-h}$ , поэтому

$$K_X(h) = \text{Mes}(X \ominus B).$$

С другой стороны, посредством эрозии и дилатации можно осуществлять фильтрацию изображений. *Условной эрозией* назовем операцию

$$Y = X \ominus B; X_1 = (X \ominus B) \cup X_1, \quad (10.9)$$

а *условной дилатацией* — операцию

$$Y = X \oplus B; X_1 = (X \oplus B) \cap X_1, \quad (10.9')$$

где  $X_1$  — некоторое множество.

Введем последовательность структурных элементов  $\{B^i\}$  и обозначим

$$X \ominus \{B^i\} = (\dots((X \ominus B^1) \ominus B^2) \dots) \ominus B^i \dots \quad (10.10)$$

последовательные эрозии и

$$X \oplus \{B^i\} = (\dots((X \oplus B^1) \oplus B^2) \dots) \oplus B^i \dots \quad (10.10')$$

последовательные дилатации множества  $X$  посредством структурных элементов  $\{B^i\}$ . *Последовательной условной эрозией* назовем операцию

$$X \ominus \{B^i\}; X_1 = [\dots[(X \ominus B^1) \cup X_1] \ominus B^2] \cup X_1 \dots], \quad (10.11)$$

а *последовательной условной дилатацией* — операцию

$$X \oplus \{B^i\}; X_1 = [\dots[(X \oplus B^1) \cap X_1] \oplus B^2] \cap X_1 \dots]. \quad (10.11')$$

Последовательность  $\{B^i\}$  может быть как конечной, так и бесконечной. Отметим, однако, не приводя доказательства, что если множество  $X_1$  ограничено, то последовательные условные операции сходятся к устойчивому результату за конечное число шагов.

Пусть  $\{B^i\}$  — бесконечная последовательность одинаковых структурных элементов, скажем, кругов радиуса  $r$  с началом в центре круга. Тогда операция  $Y = ((X \oplus B) \oplus \{B\}); X$  позволяет удалить из изображения все объекты с поперечными размерами менее  $2r$ , полностью сохранив форму оставшихся объектов. Напротив, операция  $Y = ((X \oplus B) \ominus \{B\}); X$  удаляет внутри объектов полости с поперечным размером менее  $2r$ , сохраняя при этом неизменными внешние границы объектов (рис. 10.10).

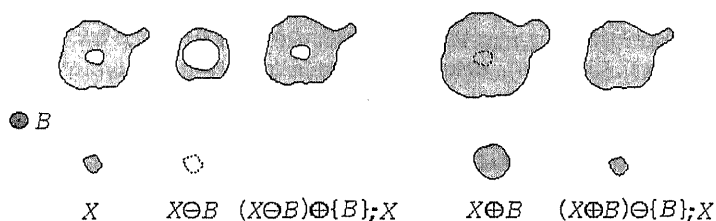


Рис. 10.10. Действие условных дилатаций (слева) и эрозии (справа) на объекты  $X$  разных размеров (вверху и внизу)

### Заполнение и пополнение

Выше мы видели, что в общем случае невозможно точно восстановить исходное множество  $X$  после эрозии  $X \ominus B$  с помощью только дилатации посредством того же структурного элемента  $B$ . Дилатация восстанавливает только часть множества  $X$ , имеющую меньше деталей, но наиболее существенную с точки зрения характеристик формы и размера. Определим операцию *заполнения* (*opening* в оригинальных работах) множества  $X$  посредством структурного элемента  $B$  как

$$Y = X_B = (X \ominus B) \oplus \bar{B}. \quad (10.12)$$

Аналогично определим операцию *пополнения* (*closing*) множества  $X$  посредством структурного элемента  $B$ :

$$Y = X^B = (X \oplus B) \ominus \tilde{B}. \quad (10.13)$$

Легко показать, что

$$(X^c)_B = (X^B)^c \text{ и } (X_B)^c = (X^c)^B. \quad (10.14)$$

В применении к изображениям эти соотношения означают, что заполнение (соответственно, пополнение) объектов и пополнение (соответственно, заполнение) фона суть операции эквивалентные.

Приведем без доказательства важное свойство этих операций — их *идемпотентность*:

$$(X_B)_B = X_B \text{ и } (X^B)^B = X^B. \quad (10.15)$$

### Применение заполнения и пополнения

Так же как эрозия и дилатация, заполнение и пополнение могут быть использованы для фильтрации изображений, сглаживания границ объектов, удаления мелких объектов и узких «хвостов» (заполнение), удаления мелких полостей и узких «каналов» (пополнение). Степень сглаживания и размеры удаляемых артефактов зависят от размеров структурного элемента, который обычно выбирается в форме круга для непрерывных изображений или правильного выпуклого многоугольника — для дискретного случая. Отметим, что при фильтрации одинаковыми структурными элементами степень искажений, вносимых в полезные детали изображения, при использовании заполнения (пополнения), оказывается значительно меньшей, чем при использовании эрозии (соответственно, дилатации). Сравните, например, на рис. 10.10 результаты операций  $X \ominus B$  и  $(X \ominus B) \oplus B$  ( $X \oplus B$  и  $(X \oplus B) \ominus B$  соответственно). Поскольку в этом примере структурный элемент симметричен относительно отражения от начала, т.е.  $\tilde{B} = B$ , то

$$(X \ominus B) \oplus B = X_B, \text{ а } (X \oplus B) \ominus B = X^B.$$

Более интересным представляется применение операции заполнения для описания формы объектов. Пусть анализируемое множество  $X$  — круг радиусом  $R$  и структурный элемент  $rB$  — круг радиусом  $r$  с началом в центре круга. Рассмотрим поведение функции

$$F(r) = \text{Mes}(X_{rB}) / \text{Mes}(X). \quad (10.16)$$



Легко понять, что до тех пор, пока радиус структурного элемента не превышает радиуса анализируемого множества,  $X_{rB} = X$ . Как только  $r$  превысит  $R$ ,  $X_{rB} = \emptyset$ , поскольку в результате эрозии, являющейся первой операцией в заполнении, будет получено пустое множество. В результате получим

$$F(r) = \begin{cases} 1, & r > R \\ 0, & r \leq R \end{cases}.$$

Пусть теперь множество  $X$  — область, ограниченная эллипсом с полуосями  $a$  и  $b$ , причем  $a > b$ . Радиус кривизны эллипса достигает своего минимального значения  $\rho_{\min} = \frac{b^2}{a}$  при пересечении с большой осью. Поэтому до тех пор, пока радиус структурного элемента будет меньше, чем  $\rho_{\min}$ , заполнение не будет приводить к изменению исходного множества и, следовательно,  $\text{Mes}(X_{rB}) = \text{Mes}(X) = \pi ab$ . С другой стороны, ясно, что как только радиус структурного элемента станет больше малой полуоси эллипса  $b$ , в результате заполнения получится пустое множество и  $\text{Mes}(X_{rB})$  примет нулевое значение. В промежутке от  $\rho_{\min}$  до  $b$   $\text{Mes}(X_{rB})$  будет монотонно убывать от  $\pi ab$  до нуля. Поэтому  $F(r)$  примет вид:

$$F(r) = \begin{cases} 1, & r \leq \rho_{\min} \\ m(r), & \rho_{\min} < r \leq b, \\ 0, & r > b \end{cases}$$

где  $m(r)$  — монотонно убывающая функция ( $0 < m(r) < 1$ ).

Иногда удобнее пользоваться функцией  $f(r) = -dF(r)/dr$ , характеризующей изменение меры анализируемого множества при заполнении его семейством монотонно увеличивающихся структурных элементов. На рис. 10.11 приведены примеры объектов разной формы и соответствующие им функции  $f(r)$ .

Функция  $f(r)$  может быть вычислена не для одиночного объекта, а, скажем, для изображения, содержащего множество объектов. Можно предполагать, что если все объекты имеют близкие размеры, то  $f(r)$  будет унимодальной, а если объекты образуют несколько групп по размерам, то в  $f(r)$  появится несколько

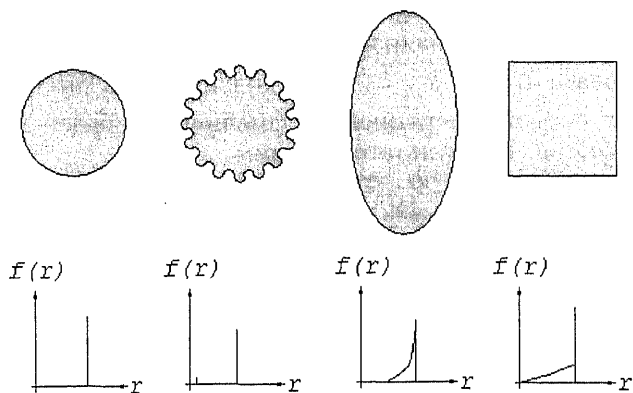


Рис. 10.11. Представление формы объектов посредством последовательных заполнений

выраженных пиков при значениях  $r$ , соответствующих этим размерам. Аналогичным образом сформировав функцию

$$G(r) = \text{Mes}(X^{rB}) / \text{Mes}(X) \quad (10.17)$$

для операции пополнения, можно использовать ее для анализа расстояний между объектами и обнаружения пространственной группировки объектов.

## 10.2. МОРФОЛОГИЧЕСКИЕ ОПЕРАЦИИ В ДИСКРЕТНОМ ПРОСТРАНСТВЕ

Обычно  $n$ -мерные дискретные данные упорядочиваются в соответствии с  $n$  целочисленными параметрами, образуя некоторую пространственную структуру. Если эти параметры изменяются регулярным образом (например, номера столбцов и строк в дискретном изображении), структура может быть представлена в виде решетки. Построим двумерную решетку следующим образом: определим в  $\mathbf{R}^2$  два линейно независимых вектора  $\mathbf{u}_1$  и  $\mathbf{u}_2$ . Решеткой назовем множество вершин всех возможных векторов вида  $\mathbf{u} = k_1\mathbf{u}_1 + k_2\mathbf{u}_2$ , где  $k_1, k_2$  — целые числа. Примеры наиболее распространенных решеток приведены на рис. 10.12.

Переход от непрерывного к дискретному пространству создает ряд проблем не только формального, но и практического характера. Принципиальная анизотропия дискретного простран-

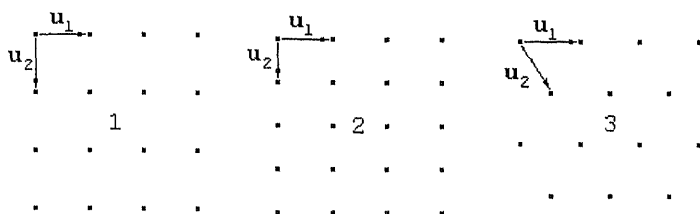


Рис. 10.12. Примеры решеток: 1-квадратная, 2-прямоугольная, 3-гексагональная

ва делает невозможным, например, поворот на произвольный угол. Возникает проблема и с нахождением расстояния, которое в непрерывном пространстве вводится достаточно естественным образом. Для некоторых типов решеток неоднозначным образом определяется понятие соседства. Последнее обстоятельство иллюстрирует рис. 10.13. Назовем множество связным, если из одной его точки к любой другой можно проложить путь, проходящий только по точкам, принадлежащим этому множеству, при этом каждая следующая точка пути должна соседствовать с текущей.

На рис. 10.13, а слева приведено три возможных задания соседства для прямоугольной решетки: соседство через стороны решетки, через узлы решетки и через стороны и узлы. Если мы примем первое определение соседства, то обнаружим, что белое поле в правой части рисунка состоит из двух частей, не связанных между собой. Следовательно, их должна разделять связная область черного цвета. Между тем такой области нет, поскольку точки черного контура тоже не связаны между собой. Если воспользуемся вторым определением соседства, получим не менее парадоксальную ситуацию: теперь точки и вне и внутри связного контура принадлежат односвязной области. Та же ситуация возникает и при третьем определении соседства.

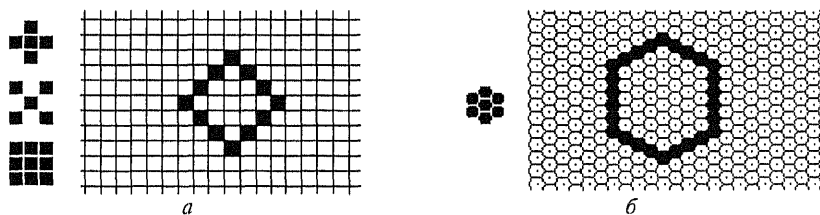


Рис. 10.13. Соседство и связность:

а — прямоугольная решетка; б — гексагональная решетка

Один из способов устранения этого противоречия состоит в том, чтобы определять по-разному соседство для белых и черных областей, скажем, для белых через стороны, а для черных — через узлы. Но тогда одни и те же операции, выполненные на изображениях, инвертированных друг относительно друга по яркости, могут приводить к различным результатам. Другой способ состоит в выборе типа решетки, не создающего вовсе этой проблемы. К такому типу относится гексагональная решетка (рис. 10.13, б). Поэтому ниже будем пользоваться этой решеткой.

Влияние анизотропии дискретного пространства демонстрирует рис. 10.14. Здесь показано поведение функции  $f(r)$ , вычисленной для объекта, представляющего дискретную аппроксимацию равностороннего треугольника на гексагональной решетке. В качестве структурного элемента используется дискретный аналог круга радиусом  $r$  — гексагон  $rH$ , где  $r$  — длина стороны гексагона (см. рис. 10.14, а слева). В первом случае (рис. 10.14, а) стороны треугольника параллельны базисным векторам решетки  $u_1$ ,  $u_2$  и вектору  $(u_1 + u_2)$ , задающему третье главное направление решетки [10.2]. Во втором случае (рис. 10.14, б) треугольник повернут на угол  $90^\circ$ .

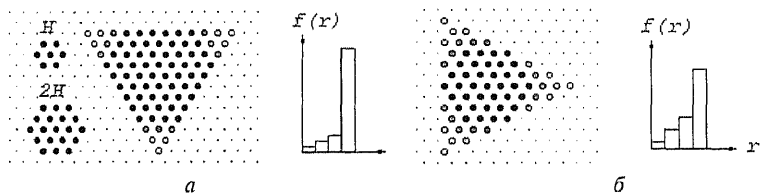


Рис. 10.14. Влияние ориентации на функцию формы объекта.

Белым обозначены точки, исчезающие на первом шаге ( $r = 1$ ); светло-серым — на втором ( $r = 2$ ); темно-серым — на третьем ( $r = 3$ ); черным — на четвертом ( $r = 4$ )

Эти особенности необходимо учитывать при реализации введенных выше морфологических операций в дискретном пространстве. Существует ряд операций, которые можно определить и в непрерывном пространстве, однако их применение имеет практический смысл только на решетках. Одна из таких операций нам уже известна. Это НМ-преобразование. НМ-преобразование, использующее различные структурные элементы, позволяет выделять особые точки на изображении. Например, точки разветвления линий на гексагональной решетке могут появляться только в конфигурациях, приведенных на рис. 10.15, причем конфигурации 1–2, 3–8 и 9–14 идентичны с точностью до

поворота вокруг центральной точки. Поэтому НМ-преобразование с использованием структурных элементов, построенных на базе конфигураций 1, 3 и 9, позволяет выявить любые точки разветвления.

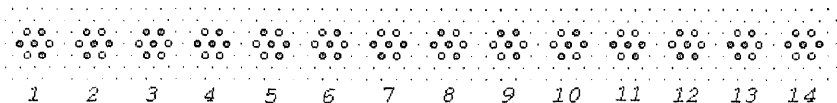


Рис. 10.15. Конфигурации, соответствующие точкам разветвления на гексагональной решетке

### Вычисление количества связных компонент

Полостями множества  $X$  называются связные компоненты множества  $X^c$ . На гексагональной решетке количество связных компонент  $n_c$  и количество полостей  $n_h$  множества  $X$  связаны соотношением

$$n_c - n_h + 1 = n \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} - n \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad (10.18)$$

где символом  $n(*)$  обозначено количество конфигураций  $*$ , встречающихся в множестве  $X$ . Доказательство этого утверждения можно найти в [10.2, р.185]. Если компоненты  $X$  не содержат полостей, то  $n_c$  просто равно их количеству, поскольку в этом случае  $X^c$  состоит из одной связной компоненты и, следовательно,  $n_h = 1$ . Но, как мы видели раньше, НМ-преобразование выделяет в исходном множестве точки, окрестность которых совпадает со структурным элементом. Используя в НМ-преобразовании структурные элементы, приведенные на рис. 10.16, получим

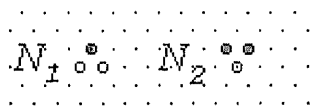


Рис.10.16. Структурные элементы, используемые для подсчета связных компонент (точкой обозначено начало)

$$\begin{aligned} n &= n_c - n_h + 1 = n \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} - n \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \\ &= \text{Mes}(X \otimes N_1) - \text{Mes}(X \otimes N_2). \end{aligned} \quad (10.19)$$

### Утончение и утолщение

Операция утончения (*thinning*) определяется как

$$Y = X \circ T = X \setminus (X \otimes T), \quad (10.20)$$

а операция *утолщения* (*thickening*) — как

$$Y = X \odot T = X \cup (X \otimes T), \quad (10.21)$$

где  $T=(T_1, T_2)$  — структурный элемент, состоящий из двух непесекающихся подмножеств  $T_1$  и  $T_2$ .

Отметим, что если начало структурного элемента принадлежит  $T_1$ , то  $X \otimes T \subseteq X$ , если же начало принадлежит  $T_2$ , то  $X \otimes T \subseteq X^c$ . Поэтому в первом случае  $Y=X \odot T=X$  при любом  $T_2$ , а во втором —  $X \odot T=X$  при любом  $T_1$ . Чтобы избежать получения этих тривиальных результатов, всегда будем полагать, что при выполнении операции утончения (соответственно, утолщения) начало структурного элемента не принадлежит  $T_2$  (соответственно,  $T_1$ ). Кроме того, можно показать, что  $(X \odot T)^c = X^c \odot T^*$ , где  $T^*=(T_2, T_1)$ . Примеры операций утончения и утолщения приведены на рис. 10.17.

Так же как и ранее, введем последовательность структурных элементов  $\{T^i\}$  и обозначим

$$X \odot \{T^i\} = (...((X \odot T^1) \odot T^2)...) \odot T^i ... \quad (10.22)$$

*последовательные утончения и*

$$X \odot \{T^i\} = (...((X \odot T^1) \odot T^2)...) \odot T^i ... \quad (10.22')$$

*последовательные утолщения* множества  $X$  с помощью последовательности структурных элементов  $\{T^i\}$ .

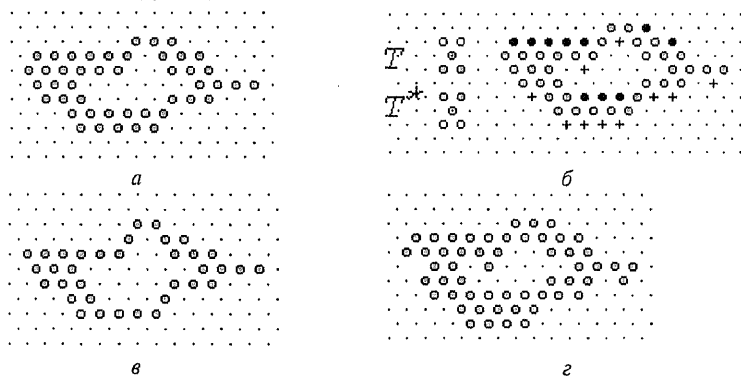


Рис. 10.17. Утончение и утолщение:

$a$  — серыми кружками помечено исходное множество;  $б$  — черными кружками помечен результат НМ-преобразования посредством структурного элемента  $T$ , а крестиками — результат НМ-преобразования посредством структурного элемента  $T^*$  (начало структурного элемента — кружок с точкой в центре);  $в$  — утончение;  $г$  — утолщение

Изучим результат последовательных утончений множества  $X$  посредством последовательности структурных элементов  $\{L^i\}$ ,  $i = 1, 6$ , где  $L^i$  отличаются друг от друга поворотом вокруг центральной точки (рис. 10.18). На крайнем правом рисунке показан установившийся результат последовательных утончений, который при последующих утончениях не изменяется.

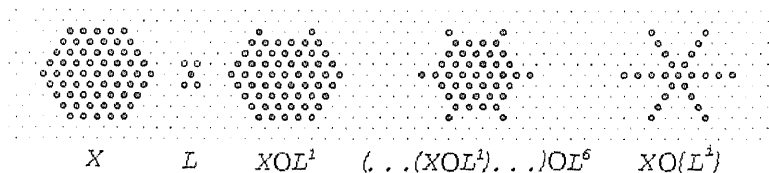


Рис. 10.18. Последовательные утончения

Приведенный пример демонстрирует применение операции утончения для построения *скелетона* (или скелета) множества  $X$ . Понятие скелетона (или скелета) достаточно интуитивно. На этом уровне его иногда пытаются описать с помощью качественной модели «степного пожара». Представим себе степной массив, покрытый сухой травой. Допустим, что одновременно вдоль всей границы массива вспыхивает огонь, распространяющийся во всех направлениях с одинаковой скоростью. В первый момент фронт распространения огня совпадает с границей. По мере его распространения различные участки фронта встречаются друг с другом и в местах встречи фронтов огонь будет гаснуть. Вот эти места самогашения огня и образуют «скелетон» массива (рис. 10.19).

Для непрерывного двумерного пространства в работе [10.2] сформулированы следующие свойства точек скелетона множества  $X$ :

- если точка  $x$  является точкой скелетона, и  $B_x$  – наибольший круг с центром в точке  $x$ , содержащийся в  $X$ , то невозможно найти содержащийся в  $X$  больший круг (не обязательно с центром в точке  $x$ ), содержащий  $B_x$ ;
- круг  $B_x$  касается границы множества  $X$  в двух или более точках.

Там же дано одно из определений скелетона: **скелетон  $S(X)$  множества  $X$  есть множество центров максимальных кругов, содержащихся в  $X$** . Под максимальным кругом подразумевается круг, касающийся границ множества  $X$  в двух или более точках. Рис. 10.20 иллюстрирует это определение.

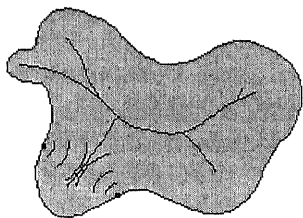


Рис. 10.19. Формирование линии гашения огня

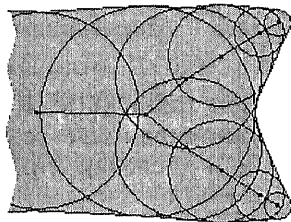


Рис. 10.20. К определению скелетона. Максимальные круги

Из этого определения (и из рис. 10.20) следует одно замечательное свойство скелетона: если каждой точке скелетона сопоставить значение радиуса максимального круга, центром которого она является, то по скелетону можно восстановить множество  $X$ , его породившее:

$$X = \bigcup_{x \in S(X)} (\rho(x)B)_x,$$

где  $\rho(x)$  – радиус максимального круга для точки  $x$  скелетона,  $B$  – круг единичного радиуса. Отметим без доказательства еще одно важное свойство скелетона: если множество  $X$  связно, то его скелетон  $S(X)$  тоже является связным множеством.

К сожалению, скелетон множества, заданного на дискретной решетке, только приближенно напоминает скелетон непрерывного множества. Более того, для одного и того же множества результат построения скелетона посредством последовательных утончений может быть различным в зависимости от порядка структурных элементов в последовательности (топологические свойства скелетона, такие как количество связных компонент, точек разветвления, ветвей, концевых точек и тому подобное при этом сохраняются). Это снова связано с анизотропией дискретного пространства. Тем не менее применение дискретного скелетона иногда оказывается чрезвычайно полезным. Так, скелетонизацию часто используют при обработке чертежей или распознавании символов для сведения линий к единичной ширине. Построение скелетона фоновой компоненты изображения, содержащего некоторое множество объектов, позволяет сегментировать его на участки, каждый из которых можно интерпретировать как зону влияния (жизненное пространство) объекта. Статистический анализ размеров, ориентации и количества соседей



таких зон применяется при анализе прочностных характеристик материалов, при исследовании поведения популяций микроорганизмов и развития лесных массивов. Множество примеров применения операций утончения, утолщения и построенной на них скелетонизации можно найти в [10.2, 10.4].

## ВОПРОСЫ К ГЛАВЕ 10

10.1. Какую операцию и какой структурный элемент можно использовать, чтобы сместить все множество  $X$  на вектор  $a$ ?

10.2. Докажите справедливость соотношений (10.14), принимая во внимание определения пополнения и заполнения и учитывая (10.5).

10.3. Предложите способ обнаружения пространственной группировки объектов с применением функции  $G(r)$ , определенной выражением (10.17).

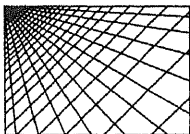
10.4. Объясните появление левого максимума в функции  $f(r)$  для второго слева объекта на рис. 10.11.

10.5. Пусть на дискретной решетке граничными точками множества  $X$  считаются точки  $x \in X$ , такие, что среди их соседей найдется хотя бы одна точка  $y \in X^c$ . Предложите способ выделения всех граничных точек.

10.6. Какие структурные элементы на гексагональной решетке позволяют выделять с помощью НМ-преобразования концевые точки линий (структурные элементы, отличающиеся только ориентацией, считать идентичными)?

10.7. Какую операцию и какой структурный элемент следует использовать для устранения изолированных черных точек на изображении, заданном на гексагональной решетке?

10.8. Допустим, что структурный элемент на гексагональной решетке может быть определен в пределах минимального гексагона (см., например, рис. 10.13, б слева). Нарисуйте все возможные структурные элементы, с которыми операция утончения не приведет к изменению связности (структурные элементы, отличающиеся только ориентацией, считать идентичными).



## ГЛАВА 11

### СЖАТИЕ ИЗОБРАЖЕНИЙ

У большинства людей понятия «код» и «кодирование» связаны с представлениями о засекречивании информации. Такое использование кодов относится к области криптографии. Однако в наше время коды нашли другое, гораздо более широкое применение в качестве экономного, удобного и надежного средства хранения и передачи информации.

Хорошо известно, что изображения, вследствие своей двумерности и многоспектральности, занимают очень большой объем памяти и их компактное хранение (архивация) предполагает сжатие исходных данных. Рассмотрим, например, дистанционное зондирование, где сжатие представляет очень серьезную проблему. По спутниковым каналам каждую секунду передается  $85 \times 10^6$  бит и типичное изображение после одного пролета состоит из  $6100 \times 6100$  пикселей в 7 спектральных каналах, что составляет 260 мегабайт данных. Поэтому алгоритмы сжатия, прежде всего, ориентированы на устранение избыточности двумерных многоспектральных данных таким образом, чтобы иметь возможность в случае необходимости полностью восстановить исходное изображение. В общем случае алгоритм сжатия данных с последующим восстановлением исходного изображения можно представить в виде блок-схемы (рис. 11.1).

Здесь первый шаг состоит в уменьшении информационной избыточности, обусловленной высокой корреляцией двумерных данных. Подавляющее большинство современных методов сжатия изображений базируется на фундаментальной идее академика А.Н. Колмогорова [11.1], заключающейся в декорреляции двумерного сигнала путем перехода от значений сигнала к слабо коррелированным переменным (трансформантам) и последующему их экономному кодированию. Здесь нужно знать особен-

ности статистической структуры изображений, поскольку, чем более коррелированы их элементы, тем большее сжатие может быть достигнуто.

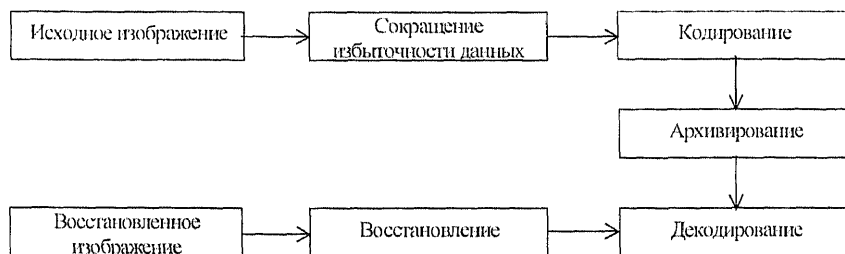


Рис. 11.1. Сжатие и восстановление данных (блок-схема)

Методы сжатия данных можно разбить на две основные группы: сжатие с сохранением информационного содержания, позволяющее полностью восстанавливать исходное изображение (сжатие без потерь), и сжатие с регулируемой степенью искажений восстанавливаемого сигнала (сжатие с потерями). В следующем разделе мы приведем ряд методов, которые не изменяют величину энтропии изображений (их информационное содержание).

### 11.1. АЛГОРИТМЫ СЖАТИЯ БЕЗ ПОТЕРЬ

Одним из старых и самых простых алгоритмов архивации без потерь является алгоритм *RLE* (*Run Length Encoding* [11.2] – групповое кодирование). Сжатие в *RLE* происходит за счет того, что в исходном изображении встречаются цепочки одинаковых байт, которые можно экономно кодировать. Различные модификации алгоритма *RLE* реализованы в огромном количестве форматов (например, в *PCX*, *GIF*, *TIFF*, *PMP* [11.3]). Следует отметить, что многие современные форматы поддерживают запись данных (не только изображений) с использованием нескольких алгоритмов архивации либо без них. Например, формат *TIFF 6.0* может сохранять изображения с использованием алгоритмов сжатия *RLE-PackBits*, *RLE-CCIT*, *LZV*, Хаффмана с фиксированной таблицей кодов, *JPEG-Lossless* [11.2] (хотя может сохранять изображение и без архивации). Здесь алгоритм *LZV* (от фамилий авторов *Lempel-Ziv-Welch*) является более сложным и универсальным алгоритмом (по отношению к *RLE*), сжимающим изображения за счет поиска, выделения и кодирования одинаковых

подцепочек в потоке байт. Алгоритм Хаффмана обеспечивает сжатие с помощью предварительного вычисления частоты появления одинаковых байт в изображении.

Говоря о лучших на сегодня алгоритмах сжатия изображений без потерь, необходимо отметить, что в 1994 году был принят стандарт *ISO/JPEG-Lossless* [11.4]. *JPEG-Lossless* ориентирован на полноцветные 24-битовые или 8-битовые в градациях яркости изображения без палитры. На восемнадцати 8-битовых тестовых изображениях нескольких категорий (космических, медицинских, фото- и компьютерно-графических, смешанных) средний уровень сжатого по *JPEG-Lossless* изображения составил 3.98 бит/точку.

Сжатие цифровых изображений без потерь качественно отличается от сжатия с регулируемой степенью искажений, поскольку абсолютная точность воспроизведения существенно ограничивает уровень сжатия за счет предварительных линейных преобразований сигнала, что обусловлено необходимостью безошибочного обратного перехода от используемых при сжатии трансформант к восстанавливаемому исходному сигналу. В практике обработки изображений зачастую нет необходимости в точном восстановлении, нужно лишь, чтобы метод сжатия не приводил к большим искажениям сигнала. В последнем случае эффективное сжатие данных обычно также дополняется количественной мерой искажений реконструированного сигнала (либо величиной среднеквадратической ошибки, либо отношением сигнал/шум и т.п.), хотя интегральные характеристики не всегда верно соответствуют субъективному качеству восстановления.

Оптимизация процедуры сжатия цифровых изображений с регулируемой степенью искажений разбивается на две части. Вначале должны быть определены статистические характеристики изображений: гистограмма уровней яркости, энтропия, корреляционные функции в различных направлениях и т.п. Лишь после этого можно подобрать метод сжатия, соответствующий свойствам изображений.

## 11.2. СВОЙСТВА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

Важнейшей мерой информационного содержания изображений является величина энтропии яркостных вариаций сигнала. Если изображение имеет  $K$  уровней яркости, а вероятность

появления уровня  $k$  равна  $P(k)$ , то энтропия  $H_e$  (без учета корреляции уровней яркости) определяется как

$$H_e = - \sum_{k=0}^K P(k) \log_2(P(k)).$$

Информационная избыточность

$$r = b - H_e,$$

где  $b$  — наименьшее число бит, которыми можно представить уровни квантования сигнала. Величину информационной избыточности изображения непосредственно вычисляют, лишь если имеется хорошая оценка энтропии, которой зачастую нет, так как соответствующие статистические свойства изображений неизвестны. Тем не менее энтропию изображений можно оценить по гистограмме уровней яркости. Пусть  $h(k)$  — частота появления уровня яркости  $k$  на исходном изображении,  $0 \leq k \leq 2^b - 1$ , и пусть размер изображения равен  $M \times N$ . Вероятность встречи на изображении уровня яркости  $k$  можно вычислить следующим образом:

$$\hat{P}(k) = \frac{h(k)}{MN},$$

тогда оценка энтропии будет

$$\hat{H}_e = - \sum_{k=0}^{2^b-1} \hat{P}(k) \log_2(\hat{P}(k)).$$

Отсюда рассчитываем информационную избыточность  $\hat{r} = b - \hat{H}_e$  и коэффициент сжатия  $C$ :

$$C = \frac{b}{\hat{H}_e}.$$

Отметим, что гистограмма дает неточную оценку энтропии вследствие корреляции уровней яркости. Более точная оценка может быть получена из гистограммы первых разностей уровней яркости.

Приведенные формулы могут быть использованы для получения теоретических пределов возможного сжатия изображений. Например, энтропия данных дистанционного зондирования со спутников находится в пределах от 4 до 5:  $H_e \in [4, 5]$ , здесь цифровой сигнал квантован на 256 уровней яркости и изображение

представлено 8 бит на пиксел. Отсюда непосредственно вычисляются границы информационной избыточности:  $\hat{r} \in [3, 4]$  бит. Это, в свою очередь, приводит к тому, что соответствующие данные можно представлять, затрачивая в среднем от 4 до 5 бит на пиксел без потери информации, и коэффициент сжатия  $C \in [1.6, 2]$ .

### 11.3. ДИСКРЕТНЫЕ ПРЕОБРАЗОВАНИЯ ИЗОБРАЖЕНИЙ В СЖАТИИ ДАННЫХ

Основная идея излагаемого подхода состоит в представлении данных изображения коэффициентами их дискретного преобразования (трансформантами). Коэффициенты преобразования упорядочиваются в соответствии с их важностью, например в соответствии с вкладом в информационное содержание, так что трансформанты с небольшим информационным содержанием можно опустить (отбросить). Оставшиеся трансформанты квантуются, кодируются и передаются или запоминаются. Важность коэффициентов можно сопоставлять, например, визуально (субъективно), в соответствии с их вкладом в яркостную картину восстанавливаемого изображения на дисплее; тем самым можно обойти вопрос корреляции элементов изображения и оценить результат сжатия. Отличной считается архивация, при которой невозможно на глаз различить первоначальное и разархивированное (восстановленное) изображения, хорошей – когда различие видно лишь для рядом находящихся отмеченных изображений. При дальнейшем увеличении степени сжатия, как правило, становятся заметны побочные эффекты, характерные для применяемого алгоритма сжатия.

Наиболее известным преобразованием, устраняющим корреляцию, является преобразование Карунена-Лоэва, привлекающее собственные векторы ковариационной матрицы  $\Sigma$  наблюдаемых данных (элементов изображения). Представим  $M \times N$  – матрицу изображения  $\mathbf{F}$  в виде вектора  $\mathbf{x} \in \mathcal{R}^{M \times N}$ , путем «лексико-графического» упорядочения (по столбцам) элементов изображения  $\mathbf{F}$ .

Пусть  $\Sigma = E(\mathbf{x} \mathbf{x}^T) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2m} \\ \cdot & \cdot & \dots & \cdot \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_{mm} \end{pmatrix}$  – ковариационная матрица

исходных данных  $\mathbf{x}$ , а  $\Phi$  – матрица, составленная из собственных векторов  $\Sigma$ .

Преобразование  $\bar{y} = \Phi^T \bar{x}$ , где  $\bar{x} = x - \bar{x}$  – центрированный вектор (изображение), переводит исходные данные в множество некоррелированных переменных. При восстановлении изображения  $\hat{F}$  исходят из частичного разложения Карунена-Лоэва путем идентификации собственных векторов, соответствующих наибольшим собственным значениям (выделение «главных компонент»  $x$ ). Пусть  $\Phi_M$  – подматрица, составленная из собственных векторов, соответствующих максимальным собственным числам. Преобразование  $x$  к главным компонентам имеет вид линейного преобразования  $y = \Phi_M^T \bar{x}$ . Это преобразование выделяет маломерное подпространство, соответствующее в базисе Карунена-Лоэва максимальным собственным значениям. Стратегия сжатия основывается здесь на привлечении трансформированных переменных с большой дисперсией (высоким информационным содержанием) и представлении изображения только этими трансформантами [11.5, гл.23.2].

Преобразование Карунена-Лоэва – единственное преобразование, гарантирующее некоррелированность сжимаемых данных, и поэтому результат их сжатия является оптимальным в статистическом смысле. Однако базисные векторы зависят от ковариационной матрицы ансамбля исходных изображений (в общем-то, неизвестной), что существенно затрудняет практическую (программную) реализацию. Поэтому преобразование Карунена-Лоэва используется лишь как ориентир при осуществлении других преобразований. Здесь можно отметить широко известные дискретные преобразования Фурье, Адамара, Уолша, косинус- и синус-преобразования, имеющие к тому же быстрые схемы реализации с вычислительной сложностью порядка  $O(MN \log_2(MN))$ . Если изображение сжимается применением дискретного преобразования, то оно обычно разбивается на блоки размером  $8 \times 8$  или  $16 \times 16$  пикселей для ускорения вычислений, поскольку тогда каждый блок преобразуется и обрабатывается отдельно. Понятно, что здесь не учитывается корреляция между блоками изображения, хотя эта корреляция может оказаться серьезным источником информационной избыточности итогового сжатия. Новым подходом, позволяющим снизить межблочную избыточность и фрагментарность преобразования, является *рекурсивное блочное* кодирование [11.4]. Приобретают популярность также различные виды дискретного вейвлет-преобразования (ДВП) [11.6].

Большая (чем у других) сравнительная близость косинус-преобразования к преобразованию Карунена-Лоэва, является одной

из причин его популярности. Возможности сжатия с применением дискретного косинус-преобразования (ДКП) показаны на рис. 11.2.

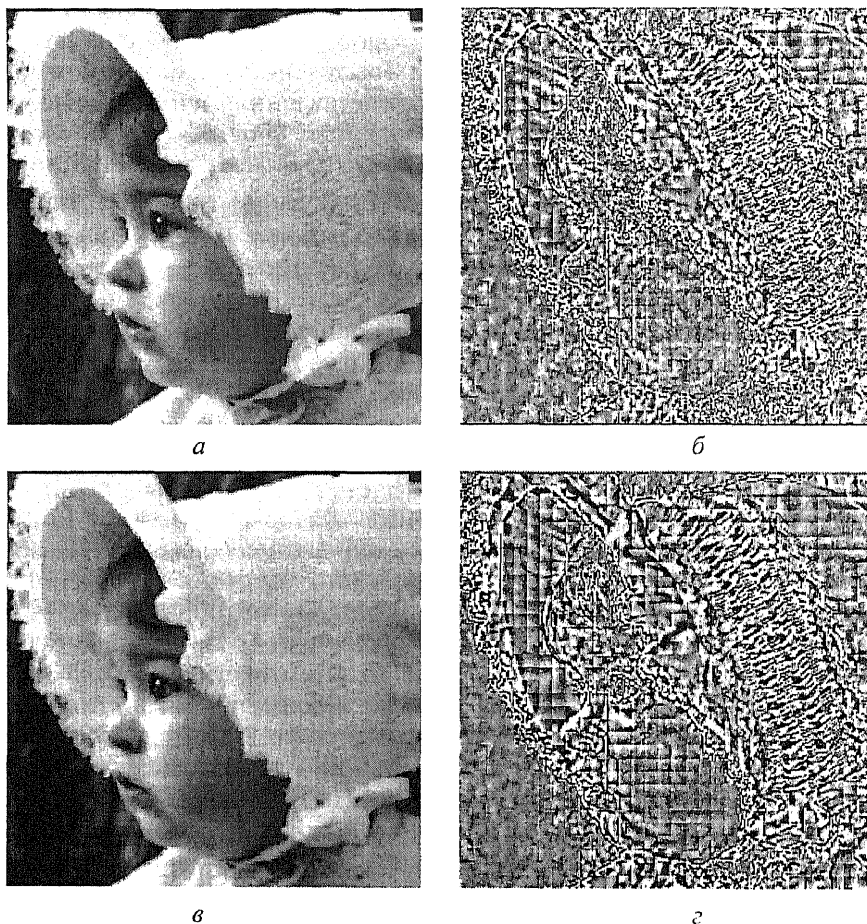


Рис. 11.2. Сжатие изображения с применением ДКП (источник – [11.4]):  
 а – восстановленное изображение, коэффициент сжатия 4.2, б – разностное изображение – разности между яркостями исходного и восстановленного изображений; в – восстановленное изображение, коэффициент сжатия 5.6, г – разностное изображение между исходным и восстановленным изображениями

ДКП в данном случае обеспечивает неплохое сжатие с небольшими вычислительными затратами, коэффициенты сжатия в представленных двух вариантах обработки равны  $C=4.2$  и



$C = 5.6$ . (В двумерном случае прямое дискретное косинус преобразование имеет вид

$$F[u, v] = \left( \frac{2^{c(u)c(v)}}{N} \right) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f[m, n] \cos \left( \frac{2m+1}{2N} u \pi \right) \cos \left( \frac{2n+1}{2N} v \pi \right),$$

$$u = 0, 1, \dots, N-1; v = 0, 1, \dots, N-1; c(k) = \begin{cases} 1/\sqrt{2}, & k = 0 \\ 1, & k \neq 0 \end{cases}.$$

Данное преобразование используется в JPEG —кодировке изображений.)

Отметим здесь, что разбиение на блоки с последующим ДКП-сжатием приводит к снижению качества восстановленного изображения при больших коэффициентах сжатия. Следовательно, сжатие изображений посредством дискретного вейвлет-преобразования заслуживает определенного внимания, поскольку оно может быть эффективно применено к целому изображению и при нем не будут присутствовать артефакты блочного разбиения [11.4].

Вейвлет-сжатие сводится к тем же этапам обработки, что и ДКП: преобразование, квантование и кодирование ДВП-трансформант. Рис. 11.3 демонстрирует восстановленные изображения с привлечением ДВП-сжатия при двух различных коэффициентах сжатия:  $C = 4.2$  и  $C = 5.6$ . Первый уровень сжатия (рис. 11.3 а, б) был достигнут обнулением 82.0 % коэффициентов ДВП, второй (большой) уровень сжатия (рис. 11.3, в, г) достигнут обнулением 87.6 % коэффициентов ДВП. Отметим отсутствие артефакта разбиения на блоки.

Вейвлеты основываются на какой-либо базисной функции  $\Phi(x)$ , называемой *материнской функцией* или *анализирующим вейвлетом*. Из нее выстраивается вейвлет-базис в виде

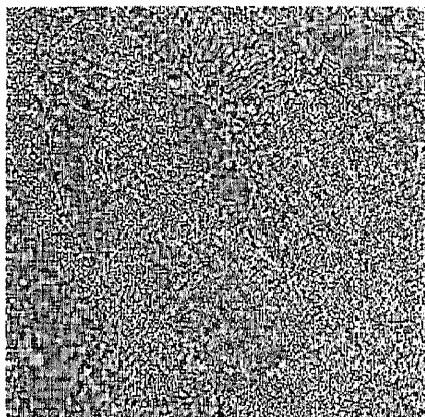
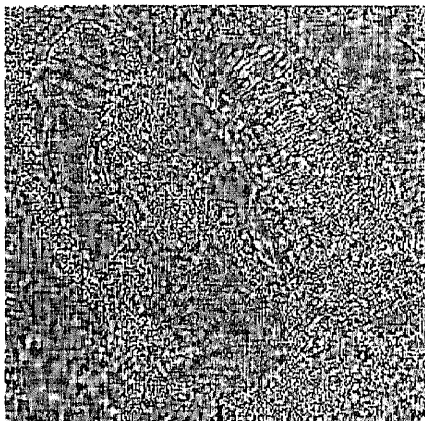
$$\Phi_{(s,l)}(x) = 2^{-(s/2)} \Phi(2^{-s} x - l).$$

Здесь масштабный множитель  $s$  (степень 2) характеризует «ширину» соответствующего вейвлета, а целочисленная переменная  $l$  — его «местоположение» на оси координат. Очевидно, что ключевым моментом в применениях является выбор материнской функции; простейшим примером является функция Хаара:

$$\Phi(x) = 1, \quad 0 \leq x < 0.5,$$

$$\Phi(x) = -1, \quad 0.5 \leq x < 1.$$

Вследствие локализованности базисных функций и по пространству и по частоте вейвлет-преобразование лучше приспособлено к анализу данных на разных масштабах (или при различном разрешении) и дает более компактное представление сигнала, нежели Фурье- или косинус-преобразования.

*a**б**в**г*

*Рис. 11.3.* Сжатие изображения с применением ДВП [12.4]:

*a* – восстановленное изображение, коэффициент сжатия 4.2; *б* – разностное изображение – разности между яркостями исходного и восстановленного изображений; *в* – восстановленное изображение, коэффициент сжатия 5.6; *г* – разностное изображение между исходным и восстановленным изображениями

## 11.4. МЕТОДЫ СЖАТИЯ НА ОСНОВЕ ПРЕДСКАЗАНИЯ

Сжатие на основе предсказания использует информационную избыточность (корреляцию данных) для формирования оценки  $\hat{f}[i, j]$  уровня яркости элемента изображения  $[i, j]$  по значениям яркости его непосредственных соседей (лежащих в окрестности  $[i, j]$ ). В той части изображения, где данные некоррелированы, оценка  $\hat{f}[i, j]$ , в общем-то, не будет совпадать с исходным значением яркости  $f[i, j]$ . Разность между оценкой и действительным значением, которую можно ожидать относительно небольшой (по абсолютной величине), далее квантуется, кодируется и передается (или запоминается) совместно с Ваметрами модели прогноза — они и будут представлять итоговый результат сжатия данных. Яркость в положении  $[i, j]$  восстанавливается по значениям вычисленной оценки  $\hat{f}[i, j]$  и разности  $d[i, j] = \hat{f}[i, j] - f[i, j]$ . Этот метод называется дифференциальной импульсно-кодовой модуляцией (ДИКМ) [11.5, гл.22.5]. Эксперименты показали, что линейный предсказатель третьего порядка (использующий три соседа — левый, верхний и диагональный) для широкого диапазона изображений достаточен для прогноза. Если изображение обрабатывается строка за строкой, то оценка  $\hat{f}[i, j]$  представляется в виде

$$\hat{f}[i, j] = a_1 f[i, j-1] + a_2 f[i-1, j-1] + a_3 f[i-1, j],$$

где  $a_1, a_2, a_3$  — Ваметры принятой модели прогноза. Эти Ваметры выбираются так, чтобы минимизировать средний квадрат ошибки оценивания  $\varepsilon = E\left[\left(\hat{f}[i, j] - f[i, j]\right)^2\right]$ . Решение, предполагающее сигнал стационарным случайным процессом с нулевым средним, использующее предсказатель третьего порядка, находится из системы уравнений

$$a_1 R[0, 0] + a_2 R[0, 1] + a_3 R[1, 1] = R[1, 0],$$

$$a_1 R[0, 1] + a_2 R[0, 0] + a_3 R[1, 0] = R[1, 1],$$

$$a_1 R[1, 1] + a_2 R[1, 0] + a_3 R[0, 0] = R[0, 1],$$

где  $R[m, n]$  является автокорреляционной функцией случайного процесса  $f[i, j]$ . Для подавляющего большинства изображений

автокорреляционная функция островершинная, имеет экспоненциальный вид и дисперсия разности  $d[i, j] = \hat{f}[i, j] - f[i, j]$  обычно существенно меньше дисперсии исходной переменной  $f[i, j]$ . Это делает возможным сжатие данных.



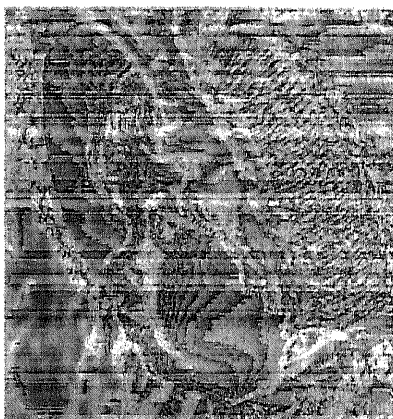
а



б



в



г

Рис. 11.4. Сжатие на основе предсказания [11.4]:

а – восстановленное изображение, коэффициент сжатия 3.8, б – разностное изображение – разности между яркостями исходного и восстановленного изображений ( $C=3.8$ ); максимальная разность – 6 уровней квантования; гистограмма изображения перед визуализацией эквализована; в – восстановленное изображение, коэффициент сжатия 6.2, г – разностное изображение между исходным и восстановленным изображениями ( $C=6.2$ ); максимальная разность – 140 уровней квантования

Очевидно, что данный алгоритм сжатия адаптивен к изображению в среднем. Он не реагирует на локальные особенности изображения, к которым в первую очередь необходимо отнести слитные участки массива, состоящие из одинаковых по яркости отсчетов. Для кодирования таких участков, как правило, применяют предсказатель нулевого порядка, в котором используется кодирование длин серий отсчетов с одинаковыми значениями [11.5, гл.22.3]. Отметим, что достаточно эффективны алгоритмы сжатия, в которых кодирование длин серий совмещается с алгоритмом линейного прогноза. Здесь необходимо дополнительно передавать также информацию о распределении длин серий. Если изображение состоит из резко отличающихся по динамике областей, то может оказаться целесообразным разбиение изображения на блоки (фрагменты).

Метод предсказания второго порядка с переменной длиной кода при кодировании разностей  $d[i, j]$  был применен для сжатия уже известного изображения и представлен на рис. 11.4; были получены коэффициенты сжатия данных  $C = 3.8$  и  $C = 6.2$ .

Отметим горизонтальные линии и ложные контуры, обусловленные методом сжатия, а также уменьшение качества восстановления при большем коэффициенте сжатия.

## 11.5. ДРУГИЕ АЛГОРИТМЫ

*Векторное квантование.* Метод сжатия посредством предварительного разбиения изображения на небольшие блоки и представления этих блоков в виде векторов. Такие векторы кодируются с использованием уникальных кодовых слов из словаря этих слов, так что передаче (или хранению) подлежат не сами векторы, а их коды. Очевидно, что словарь кодов (или кодовая книга) должен передаваться вместе с кодовыми последовательностями.

*Фрактальный алгоритм.* Фрактальная архивация основана на представлении изображения в компактной форме с помощью коэффициентов системы итерируемых функций (Iterated Function System – IFS [11.7]). IFS представляет собой набор трехмерных аффинных преобразований, переводящих одно изображение в другое (машина Барнсли). В процессе итераций получается изображение, которое перестает изменяться. Это изображение называется «неподвижной точкой» (или аттрактором) данной IFS. Соответствующая теория гарантирует наличие ровно одной неподвижной точки для каждой IFS. Более того, система итери-

руемых функций задает фрактал (в определенном смысле самоподобный математический объект). Фактически фрактальное сжатие — это поиск самоподобных областей в изображении и определение для них Ваметров аффинных преобразований, что требует чрезвычайно много времени. Поэтому исследования сейчас направлены на решение данной проблемы.

Достаточно давно велась работа по стандартизации сжатия изображений. The Joint Photographic Experts Group (*JPEG*) — подразделение в рамках ISO — Международной организации по стандартизации разработало международный стандарт для полноцветных (24-битовых) изображений [11.8]. Этот стандарт — *JPEG* — один из мощных алгоритмов сжатия. Для него реализованы четыре варианта сжатия:

- последовательное, основанное на ДКП, сжатие;
- прогрессивный вариант ДКП-сжатия, использующий пирамидальное представление изображений;
- последовательное, основанное на предсказании сжатие без потерь (*JPEG-Lossless*);
- иерархическое сжатие, соответственно с потерями и без потерь.

Оперирует алгоритм областями  $8 \times 8$ , на которых яркость и цвет меняются сравнительно плавно. Вследствие этого при разложении матрицы такой области в двойной ряд по косинусам значимыми оказываются только первые коэффициенты. Кроме того, система цветового восприятия человека слабо распознает определенные частоты (цветовое изменение картинки). Поэтому можно аппроксимировать некоторые коэффициенты более грубо без заметной потери качества изображения, используя для этого квантование коэффициентов. При таком преобразовании теряется часть информации, но могут достигаться большие коэффициенты сжатия.

Рассмотрим поэтапно основные процедуры алгоритма, считая, что сжимается 24-битовое изображение [11.2].

*Этап 1.* Переводим изображение из цветового пространства RGP с компонентами, отвечающими за красную (Red), зеленую (Green) и синюю (Blue) составляющие цвета пиксела, в цветовое пространство YCrCb (иногда называют YUV). В нем Y — яркостная составляющая, а Cr, Cb — компоненты, отвечающие за цвет (хроматический красный и хроматический синий). За счет того, что человеческий глаз менее чувствителен к вариациям цветовой компоненты, чем к яркости, появляется возможность архивировать массивы для Cr и Cb составляющих с большими потерями и

соответственно большими коэффициентами сжатия. Подобное преобразование уже давно используется в телевидении.

Упрощенно перевод из цветового пространства RGP в цветное пространство YCbCr можно представить так:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.5 & -0.4187 & -0.0813 \\ 0.1687 & -0.3313 & 0.5 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix}.$$

Обратное преобразование осуществляется умножением вектора YUV на обратную матрицу.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.34414 & -0.71414 \\ 1 & 1.772 & 0 \end{bmatrix} \times \begin{bmatrix} Y \\ U \\ V \end{bmatrix} - \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix}.$$

*Этап 2.* Разбиваем исходное изображение на матрицы  $8 \times 8$ . Формируем из каждой три рабочие матрицы ДКП — по 8 бит отдельно для каждой компоненты. При больших коэффициентах сжатия этот шаг может выполняться чуть сложнее.

*Этап 3.* Применяем ДКП к каждой рабочей матрице. При этом получаем матрицу, в которой коэффициенты в левом верхнем углу соответствуют низкочастотной составляющей изображения, а в правом верхнем — высокочастотной.

*Этап 4.* Производим квантование коэффициентов. В принципе это просто деление рабочей матрицы на матрицу квантования поэлементно. Для каждой компоненты ( $Y$ ,  $U$ , и  $V$ ), в общем случае задаем свою матрицу квантования (МК),  $q[u, v]$  например для  $Y$ :

$$Yq[u, v] = \text{IntegerRound} \left( \frac{Y[u, v]}{q[u, v]} \right).$$

На этом шаге осуществляется управление степенью сжатия и происходят самые большие потери. Понятно, что, задавая МК с большими коэффициентами, получаем больше нулей и, следовательно, большую степень сжатия. В стандарт *JPEG* включены рекомендованные разработчиками МК, построенные опытным путем. Матрицы для большего или меньшего коэффициентов сжатия получают путем умножения исходной матрицы на некоторое (рекомендованное) число *gamma*.

С квантованием связаны и специфические эффекты алгоритма. При больших значениях числа *gamma* потери в низких час-

тотах могут быть настолько велики, что восстановленное изображение распадется на квадраты  $8 \times 8$ . Потери в высоких частотах могут проявиться в так называемом эффекте Гиббса, когда вокруг контуров с резким переходом цвета образуется своеобразный «нимб».

*Этап 5.* Переводим матрицу квантованных коэффициентов  $8 \times 8$  в 64-элементный вектор при помощи «зигзаг»-сканирования, когда берутся элементы с индексами  $[0,0]$ ,  $[0,1]$ ,  $[1,0]$ ,  $[2,0]$ ,  $[1,1]$ ,  $[0,2]$ ,  $[0,3]$ ,  $[1,2]$ ,... В результате, в начале вектора получаем трансформанты, соответствующие низким частотам, а в конце — высоким.

*Этап 6.* Свертываем вектор с помощью алгоритма группового кодирования (*RLE*). При этом получаем *Вы* типа (*пропустить*, *число*), где *пропустить* является счетчиком пропускаемых нулей, а *число* — значением, которое необходимо поставить в следующую ячейку. Так, вектор 42 3 0 0 0 -2 0 0 0 0 1 ... будет свернут в *Вы* (0,42) (0,3) (3, -2) (4,1).

*Этап 7.* Свертываем получившиеся *Вы* кодированием по Хаффману с фиксированной кодовой таблицей.

Процесс восстановления изображения в этом алгоритме полностью симметричен. Метод позволяет сжимать некоторые изображения в 10...15 раз без серьезных потерь.

## ВОПРОСЫ К ГЛАВЕ 11

11.1. Какие свойства изображений обуславливают возможность их сжатия, и как осуществляются сжатие и восстановление информации (блок-схема)?

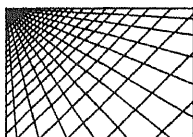
11.2. Что подразумевается под информационной избыточностью изображений и коэффициентом их сжатия?

11.3. Охарактеризуйте статистически оптимальный метод сжатия изображений?

11.4. В чем состоят идеи основных методов сжатия: на основе дискретных преобразований, на основе предсказания, фрактального алгоритма?

11.5. Опишите основные этапы алгоритма JPEG; на каком из них происходит сжатие информации и за счет чего?





## ГЛАВА 12

### КОМПЬЮТЕРНАЯ ТОМОГРАФИЯ

Желание заглянуть внутрь непрозрачного объекта, не разрушив его, существовало на протяжении многих веков развития человечества. Первым шагом в решении этой проблемы было открытие Рентгеном незадолго до конца 1895 г. X-лучей, проникающих через плотные вещества. Это открытие произвело ошеломляющее впечатление не только на ученых того времени, но и на всех образованных людей мира. Ведь X-лучи, которые впоследствии были названы рентгеновскими, позволяли заглянуть внутрь непрозрачных тел и видеть сквозь них. Естественно, что самый большой интерес к практическому применению рентгеновских лучей проявила медицина. Рентгеновские лучи позволяли получать изображения внутренних органов человека, обнаруживать посторонние предметы внутри его тела, переломы и т.п.

В основе формирования рентгеновских изображений лежит использование эффекта неодинаковой рентгеновской плотности веществ. Одни вещества пропускают лучи лучше, другие хуже. Пройдя через тело и попав на чувствительную пленку, лучи засвечивают участки пленки тем сильнее, чем меньше плотность вещества.

Возможность оценки взаимного расположения различных органов тела, их точной геометрической формы при таком методе исследования существенно ограничена. Основной причиной является то, что мы получаем плоское (двумерное) теневое изображение объемного (трехмерного) объекта. Теневое рентгеновское изображение представляет собой сумму изображений слоев тела, которые находятся на различных расстояниях от пленки. При этом внутренние органы тела на рентгеновском изображении наслаиваются друг на друга и важные особенности их пространственного расположения значительно искажаются или

полностью утрачиваются. Задачи получения изображения каждого изолированного слоя объекта, не искаженного никакими «наложениями», и восстановления его внутренней структуры решает современная компьютерная томография (от греческого *tomos* — слой, срез).

Математические основы компьютерной томографии были заложены задолго до появления первых рентгеновских компьютерных томографов. Еще в 1917 г. математик Радон предложил метод решения обратной задачи интегральной геометрии, состоящий в восстановлении (реконструкции) многомерных функций по их интегральным характеристикам. Однако этот метод не находил практического применения до тех пор, пока не появились, во-первых, рентгеновские установки, позволяющие получать большое количество высококачественных снимков, необходимых для восстановления внутренней структуры реальных объектов, во-вторых, быстродействующие ЭВМ, способные эти снимки обрабатывать. Первый в мире рентгеновский компьютерный томограф был продемонстрирован Хаунсфилдом в 1972 г. Внедрение методов компьютерной томографии в медицину позволило существенно повысить эффективность диагностики и обеспечило создание новых методов лечения. В настоящее время методы компьютерной томографии также широко используются в электронной и рентгеновской микроскопии — для получения структур кристаллов и макромолекул; в геофизике — для поиска и разведки месторождений полезных ископаемых; в астрофизике — для исследования полей планет и в других областях науки и техники.

### 12.1. ПОЛУЧЕНИЕ ПРОЕКЦИЙ

В основе большинства томографов лежит идея, состоящая в том, что внутреннюю структуру объекта можно представить, получив ряд параллельных поперечных сечений. Поэтому главная задача компьютерной томографии — получить двумерное (плоское) изображение поперечного сечения исследуемого объекта, которая и будет рассмотрена далее.

Метод получения двумерного томографического изображения содержит два этапа. На первом этапе формируются проекционные данные, на втором — по проекционным данным восстанавливается изображение поперечного сечения.

Чтобы определить внутреннюю структуру объекта, необходимо получить информацию о ней. Для этого используют излучение, проникающее сквозь объект. Современные медицинские

компьютерные томографы имеют различные типы аппаратной реализации сканирования объектов. В основном это системы с веерным или параллельным пучком. В первом случае лучи расходятся от источника рентгеновского излучения в одной плоскости, но под различными углами, образуя веер. Детекторы излучения располагаются на дуге с другой стороны исследуемого объекта. Этот метод сканирования позволяет регистрировать проекционные данные с большой скоростью, что существенно уменьшает искажения, вызванные движением пациента. Для реализации второго метода необходима линейка излучателей, формирующая параллельные лучи также в одной плоскости. Стоимость такой системы весьма высока. Кроме того, надо проводить калибровку линейки излучателей с целью уменьшения яркостных искажений получаемых томограмм. Однако алгоритмы восстановления томографических изображений по проекционным данным для параллельного пучка проще, чем для веерного. Поэтому далее мы ограничимся рассмотрением именно этих алгоритмов. Подробное описание других методов сканирования и восстановления приведено в работах [12.1, 12.2].

Пусть нам необходимо определить плотность распределения вещества  $u(x, y)$  в сечении объекта. Исследуемый объект в пределах тонкого поперечного слоя просвечивается параллельным пучком хорошо сфокусированных рентгеновских лучей (рис. 12.1.). Направление лучей составляет некоторый угол  $\theta$  с осью  $x$ . Лучи ослабляются веществом, находящимся внутри объекта, пропорционально его плотности. С противоположной стороны объекта располагается устройство (линейка детекторов), регистрирующее интенсивность каждого луча, прошедшего через объект. При этом полагается, что лучи распространяются в объекте вдоль прямой линии  $l$ , определяемой уравнением

$$x \cos \theta + y \sin \theta - s = 0, \quad (12.1)$$

где  $s$  — расстояние от начала координат до соответствующего луча (рис. 12.1). Тогда интенсивность луча на выходе из объекта равна интегралу от искомой функции  $u(x, y)$  вдоль траектории луча  $l$ :

$$p(s, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) \delta(x \cos \theta + y \sin \theta - s) dx dy, \quad (12.2)$$

где траектория  $l$  описывается уравнением (12.1);  $\delta(\bullet)$  — дельта-функция. Регистрируемое излучение  $p(s, \theta)$  называется радонов-

ским образом, а преобразование (12.2) — преобразованием Радона. Далее радоновский образ  $p(s, \theta)$ , вычисленный для фиксированного угла  $\theta$ , будем называть проекцией.

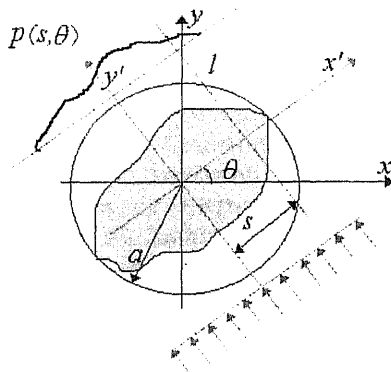


Рис. 12.1. Схема получения проекций

Связь между координатами  $(x, y)$  исходной и  $(x', y')$  повернутой на угол  $\theta$  прямоугольными системами координат определяется соотношениями [12.3]:

$$\begin{cases} x' = x \cos \theta + y \sin \theta; \\ y' = -x \sin \theta + y \cos \theta; \end{cases} \quad \begin{cases} x = x' \cos \theta - y' \sin \theta; \\ y = x' \sin \theta + y' \cos \theta. \end{cases} \quad (12.3)$$

Очевидно, что в повернутой на угол  $\theta$  системе координат уравнение прямой (12.1) и соотношение (12.2) имеют вид:

$$x' = s, \quad (12.4)$$

$$p(s, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x' \cos \theta - y' \sin \theta, x' \sin \theta + y' \cos \theta) \delta(x' - s) dx' dy'. \quad (12.5)$$

Следует отметить, что при замене переменных в двойных интегралах, т.е. при переходе от переменных  $(x, y)$  к переменным  $(x', y')$ , связанных однозначным функциональным преобразованием

$$\begin{cases} x = \psi_1(x', y'); \\ y = \psi_2(x', y'), \end{cases}$$

подынтегральная функция домножается на модуль якобиана преобразования [12.4]:

$$J = \det \begin{bmatrix} \frac{\partial x}{\partial x'} & \frac{\partial x}{\partial y'} \\ \frac{\partial y}{\partial x'} & \frac{\partial y}{\partial y'} \end{bmatrix}.$$

Очевидно, что якобиан преобразования (12.3) равен 1.

Двумерный интеграл в (12.5) можно свести к одномерному, воспользовавшись фильтрующим свойством дельта-функции:

$$p(s, \theta) = \int_{-\infty}^{\infty} u(s \cos \theta - y' \sin \theta, s \sin \theta + y' \cos \theta) dy'. \quad (12.6)$$

Проекции  $p(s, \theta)$  вычисляются под всевозможными углами  $\theta$  и для тех значений  $s$ , при которых двумерная функция  $u(x, y)$  отлична от нуля. На практике величина  $s$  ограничивается физическими размерами исследуемого объекта, а угол  $\theta$  изменяется в пределах от  $0^\circ$  до  $180^\circ$ , так как при изменении угла на  $180^\circ$  просвечивание ведется в строго обратном направлении, поэтому

$$p(s, \theta) = p(-s, \theta + \pi). \quad (12.7)$$

Удобно ввести в рассмотрение окружность радиусом  $a$ , охватывающую исследуемое поперечное сечение. В этом случае интеграл (12.6) равен:

$$p(s, \theta) = \int_{-\sqrt{a^2 - s^2}}^{\sqrt{a^2 - s^2}} u(s \cos \theta - y' \sin \theta, s \sin \theta + y' \cos \theta) dy', \quad (12.8)$$

где  $0 \leq \theta < \pi$ .

Таким образом, каждое значение радоновского образа  $p(s, \theta)$  есть интеграл от тех значений функции  $u(x, y)$ , которые она принимает вдоль луча  $l$ , определяемого параметрами  $-a \leq s \leq a$  и  $0 \leq \theta < \pi$ .

Важно отметить, что аргументы  $(s, \theta)$  радоновского образа существенно отличаются от полярной системы координат  $(\rho, \varphi)$ , которая связана с прямоугольной системой координат  $(x, y)$  следующими выражениями:

$$\begin{cases} \rho = \sqrt{x^2 + y^2}; \\ \varphi = \arctg \frac{y}{x}, \end{cases} \quad \begin{cases} x = \rho \cos \varphi; \\ y = \rho \sin \varphi. \end{cases} \quad (12.9)$$

Для полярной системы координат, так же как и для системы координат  $(s, \theta)$ , справедливо, что функция двух переменных  $u(\rho, \varphi)$  удовлетворяет условию

$$u(\rho, \varphi) = u(-\rho, \varphi + \pi), \quad (12.10)$$

аналогичному (12.7) для радоновских образов.

При задании двумерной функции  $u(\cdot)$  в полярных координатах  $(\rho, \varphi)$  мы определяем ее значение для любой пары вещественных чисел, которые являются координатами точки на плоскости. Для полярных переменных выполняется условие  $u(0, \varphi_1) = u(0, \varphi_2)$  при любых значениях углов  $\varphi_1$  и  $\varphi_2$ , так как, во-первых, точка  $(0, \varphi)$  соответствует началу координат, во-вторых, функция  $u(\cdot)$ , описывающая некоторую физическую величину, может иметь только одно значение в заданной точке на плоскости. Напротив, в общем случае для радоновских образов:  $p(0, \theta_1) \neq p(0, \theta_2)$  при  $\theta_1 \neq \theta_2$ , так как радоновские образы  $p(0, \theta_1)$  и  $p(0, \theta_2)$  представляют собой интегралы, вычисленные вдоль прямых, проходящих через начало координат под различными углами  $\theta_1$  и  $\theta_2$ . Поэтому пара вещественных чисел  $(s, \theta)$  из области определения радоновских образов не может быть интерпретирована как точка на плоскости в полярных координатах.

Определим связь между пространствами  $(s, \theta)$ ,  $(\rho, \varphi)$  и  $(x, y)$ . С учетом (12.9) уравнение прямой (12.1) в полярных координатах имеет вид

$$\rho \cos \varphi \cos \theta + \rho \sin \varphi \sin \theta - s = 0. \quad (12.11)$$

Из (12.11) следует: если луч выходит из начала координат под углом  $\theta$  к оси  $x$ , то прямая  $l$ , перпендикулярная этому лучу и проходящая через точку с координатами  $(\rho, \varphi)$  (или  $(x, y)$ ), находится на расстоянии

$$s = \rho \cos(\theta - \varphi) = x \cos \theta + y \sin \theta \quad (12.12)$$

от начала координат (см. рис. 12.2.а). Таким образом, кривая (12.12) в пространстве  $(s, \theta)$  является геометрическим местом

точек, которому соответствуют прямые, проходящие через точку с координатами  $(\rho, \varphi)$  (или  $(x, y)$ ). На рис. 12.2, б приведена эта кривая (12.11) для прямых, проходящих через точку с координатами  $(x = 5, y = 1) \equiv (\rho = 5.099, \varphi = 0.197)$ .

В компьютерных томографах проекции вычисляются для конечного числа пар  $(s, \theta)$ , так как в реальных устройствах технически невозможно получить бесконечное число проекций и измерять интенсивность излучения для всех возможных значений  $s$ . Поэтому проекционные данные  $p(s, \theta)$  в практических приложениях представляют собой дискретную функцию двух переменных  $(s, \theta)$ . Далее мы будем рассматривать случай, когда отсчеты функции берутся с равномерным шагом  $\Delta s$  и  $\Delta \theta$  по соответствующим переменным  $s$  и  $\theta$ . Положим, что источники излучения образуют параллельный пучок лучей. Число источников и соответственно число лучей равно  $N_s = [2a/\Delta s]$ , а число проекций —  $N_\theta = [\pi/\Delta \theta]$ , где квадратные скобки  $[\cdot]$  обозначают операцию взятия целой части вещественного числа. Таким образом, проекционные данные на практике представляют собой двумерный массив размером  $N_\theta \times N_s$ , заданный на прямоугольной сетке, где  $N_\theta$  — число строк,  $N_s$  — число столбцов (рис. 12.3). Томограмма восстанавливается также для дискретных значений пространственных координат  $(x, y)$ . Обычно томограмма вычисляется на прямоугольной равномерной сетке отсчетов, образующих строки и столбцы, причем  $N_x = N_y = [2a/\Delta]$ , где  $\Delta$  — шаг дискретизации, одинаковый для пространственных переменных  $x$  и  $y$ . Проблема выбора параметров  $N_s$ ,  $N_\theta$ ,  $N_x$  и  $N_y$  будет рассмотрена в разделе 12.5.

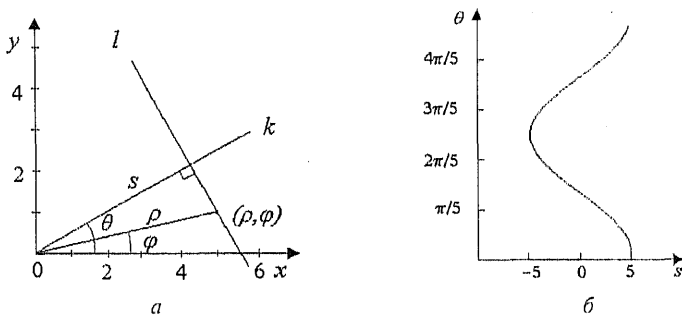


Рис. 12.2. Связь между  $(\rho, \varphi)$  и  $(s, \theta)$ -пространствами

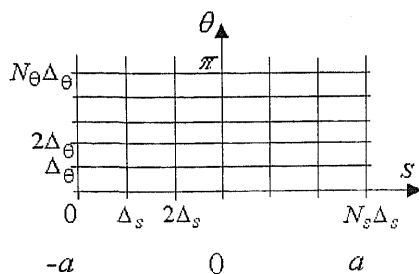


Рис. 12.3. Прямоугольная сетка в пространстве  $(s, \theta)$

В качестве примера вычислим радоновский образ для двух гауссовских импульсов, описываемых соотношением

$$u(x, y) = \sum_{i=1}^2 \exp\left\{-\frac{(x-c_{xi})^2 + (y-c_{yi})^2}{2b^2}\right\}. \quad (12.13)$$

На изображении эти импульсы выглядят как окружности, пространственное положение которых задается параметрами  $(c_{x1}, c_{y1})$  и  $(c_{x2}, c_{y2})$ , а радиус — параметром  $b$ . Подставляя (12.13) в (12.6), находим [12.5]:

$$p(s, \theta) = \sum_{i=1}^2 b\sqrt{2\pi} \exp\left\{-\frac{(c_{xi} \cos \theta + c_{yi} \sin \theta - s)^2}{2b^2}\right\}.$$

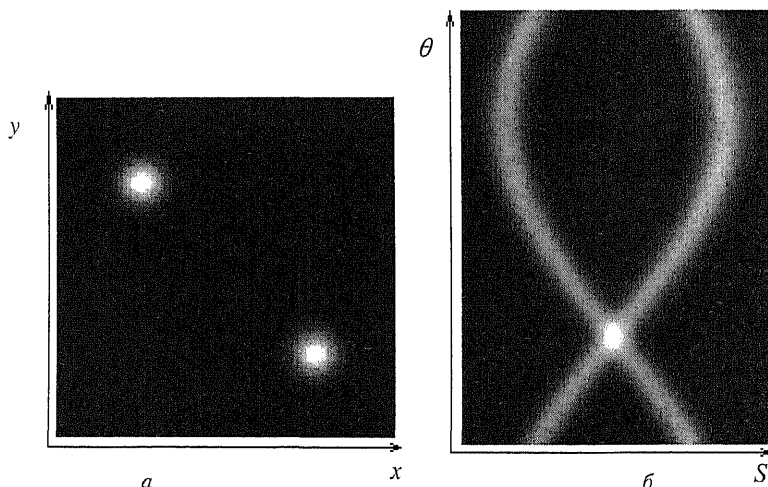


Рис. 12.4. Функция (а) и ее радоновский образ (б)



Функция (12.13) и соответствующий ей радоновский образ изображены на рис. 12.4. На рис. 12.4, б число лучей равно 128, а число проекций — 180. Очевидно, что кривые на рис. 12.4, б., представляющие собой радоновские образы гауссовских импульсов, подобны кривой (12.12). Из рис. 12.4 также следует, что функция и радоновский образ совсем не похожи друг на друга. Однако между радоновским образом и функцией, порождающей его, имеется взаимно однозначное соответствие, которое и лежит в основе всех алгоритмов реконструкции томографических изображений.

## 12.2. КЛАССИЧЕСКАЯ ТОМОГРАФИЯ

До появления ЭВМ в медицине использовалась так называемая классическая томография. Ее идея состоит в следующем. Пусть нам необходимо получить изображение объекта в плоскости  $B$  (рис. 12.5). Для этого фотопленку помещают в плоскость  $C$ , а источник рентгеновского излучения в плоскость  $A$ . Плоскости  $A$  и  $C$  параллельны плоскости  $B$ . Источник рентгеновского излучения и фотопленка перемещаются в противоположных направлениях с одинаковой скоростью. В этом случае точка пересечения осей источника рентгеновского излучения лежит на плоскости  $B$ . Потому изображение плоскости  $B$ , в частности точек  $a$  и  $b$  (см. рис. 12.5), на фотопленке в плоскости  $C$  будет неподвижным. В то же время точки, которые лежат вне плоскости  $B$ , будут отображаться в различные места фотопленки на плоскости  $C$ . Поэтому на фотопленке изображение плоскости  $B$  отображается четко, а изображения остальных сечений объекта «размазываются» за счет движения, создавая искажения томографического изображения. Несмотря на относительную простоту описанного метода, это было лишь частичное решение задачи формирования томографического изображения сечения, так как получаемое классическим методом изображение сечения  $B$  остается значительно затененным другими слоями исследуемого объекта.

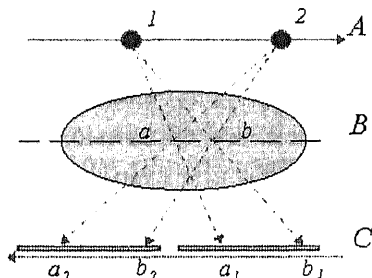


Рис. 12.5. Схема классической томографии

### 12.3. АЛГОРИТМ ОБРАТНОГО ПРОЕЦИРОВАНИЯ

Простейшим алгоритмом реконструкции изображений в компьютерной томографии является алгоритм обратного проецирования, в соответствии с которым оценка плотности  $u(x, y)$  вычисляется следующим образом. Проекция  $p(s, \theta)$  функции двух переменных  $u(x, y)$  для каждого значения угла  $\theta$  представляет собой одномерную функцию. Ее можно преобразовать в двумерную (рис. 12.6), зафиксировав угол  $\theta$  и растянув (выполнив обратное проецирование) по всей плоскости  $(x, y)$  в соответствии с выражением

$$p_{\theta}(x, y) = p(x \cos \theta + y \sin \theta, \theta). \quad (12.14)$$

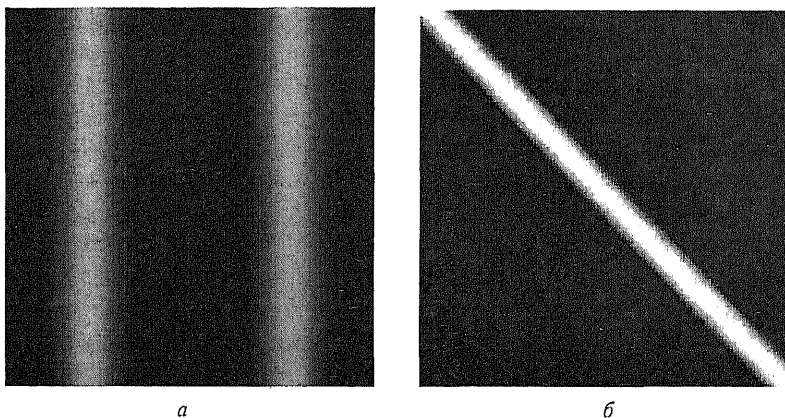


Рис. 12.6. «Растянутые» проекции функции (12.13) при  $\theta = 0^\circ$  (а) и  $\theta = 45^\circ$  (б)

Очевидно, что сечение двумерной функции  $p_{\theta}(x, y)$  плоскостью, перпендикулярной плоскости  $(x, y)$  и проекция которой на плоскость  $(x, y)$  с осью  $x$  составляет угол  $\theta$ , равно  $p(s, \theta)$ . Далее складываем все обратные проекции  $p_{\theta}(x, y)$  для  $0 \leq \theta < \pi$ . В результате получим суммарное изображение  $u_{\Sigma}(x, y)$ , которое используется в качестве оценки функции плотности  $u(x, y)$ . Суммарное изображение определяется соотношением

$$u_{\Sigma}(x, y) = \int_0^{\pi} p(x \cos \theta + y \sin \theta, \theta) d\theta. \quad (12.15)$$

При восстановлении томограмм методом обратного проектирования по дискретным проекционным данным необходимо использовать интерполяцию, так как линия, вдоль которой необходимо вычислить интеграл (12.15), чтобы найти оценку  $u_{\Sigma}(x_k, y_l)$  для дискретных значений координат  $(x_k, y_l)$ , определяется уравнением (12.12). На рис. 12.7 показаны эта

линия и прямоугольная сетка, в узлах которой известны проекционные данные, полученные с помощью равномерно распределенных параллельных лучей. Очевидно, что линия, вдоль которой вычисляется интеграл, не проходит через узлы сетки. Обычно используют метод интерполяции по ближайшему значению, при этом интеграл (12.15) заменяется на сумму

$$u_{\Sigma}(x_k, y_l) = \Delta_{\theta} \sum_{j=0}^{N_{\theta}-1} p(i\Delta s, j\Delta\theta), \quad (12.16)$$

где  $i$  выбирается из условия минимума значения выражения  $|i\Delta s - x_k \cos j\Delta\theta + y_l \sin j\Delta\theta|$ . Выражение под знаком абсолютной величины является дискретным аналогом уравнения прямой (12.1).

Операция обратного проектирования имеет простую геометрическую интерпретацию. На рис. 12.8, а показана схема получения трех проекций под углами  $\theta_1$ ,  $\theta_2$  и  $\theta_3$  по двумерному изображению  $u(x, y)$ , которое описывается функцией (12.13). Полученные проекции  $p(s, \theta_j)$ ,  $j = \overline{1, 3}$  растягиваем в соответствии с (12.14) и суммируем. Результат реконструкции функции  $u(x, y)$  по трем проекциям показан на рис. 12.8, б. Откуда видно, что, несмотря на искажения в виде полос, изображение, восстановленное лишь по трем проекциям, имеет много общего с функцией  $u(x, y)$ . Полосы являются результатом «растягивания» проекций. По их направлениям можно оценить углы проекций. При обратном проектировании каждая точка на изображении превращается в многолучевую звезду, число лучей которой равно удвоенному числу проекций. С увеличением числа проекций эти лучи будут сливаться и восстанавливаемое изображение все больше будет похоже на функцию  $u(x, y)$ , однако оно с ней никогда не совпадет.

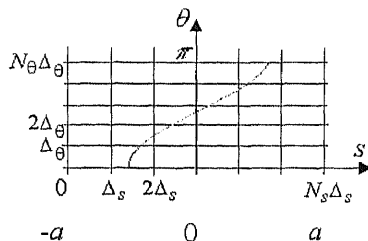


Рис. 12.7. Линия, вдоль которой вычисляется интеграл

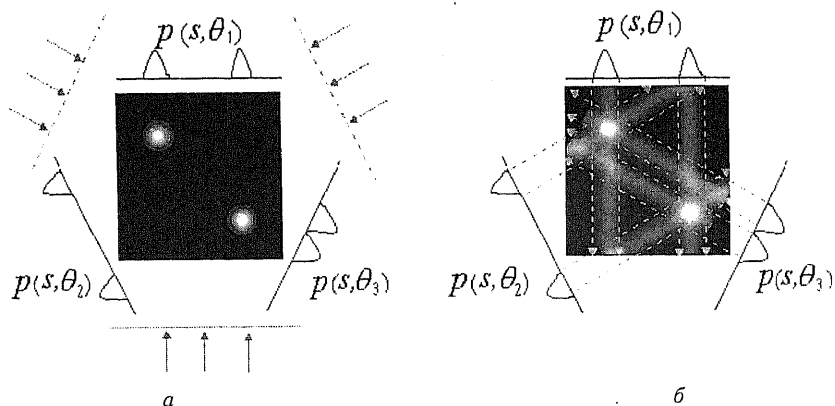


Рис. 12.8. Схема восстановления томограммы по алгоритму обратного проецирования  
 а – получение проекций; б – суммирование обратных проекций

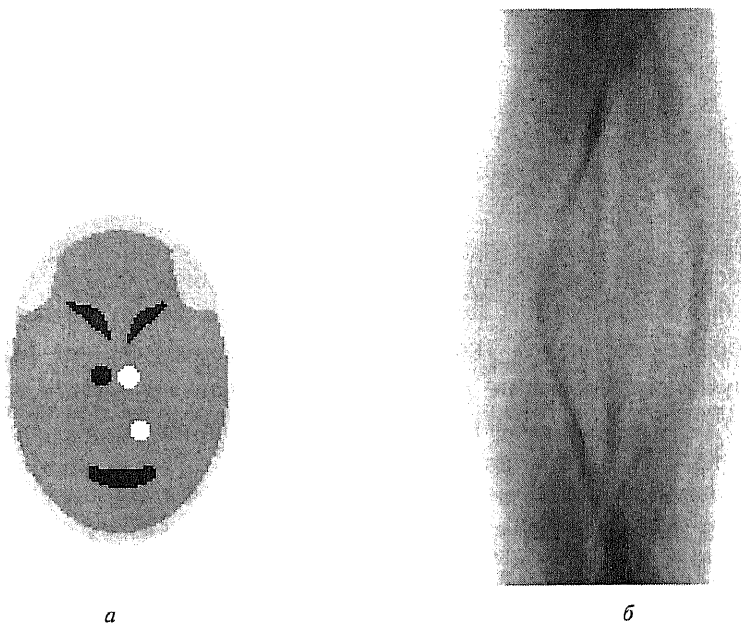


Рис. 12.9. Пример восстановления томографического изображения методом обратного проецирования.  
 а – эталонное изображение  $u(x, y)$ ; б – проекции  $R(s, \theta)$ ; в...д – результаты восстановления алгоритмом обратного проецирования по 15, 45 и 180 проекциям (см. также с. 319).

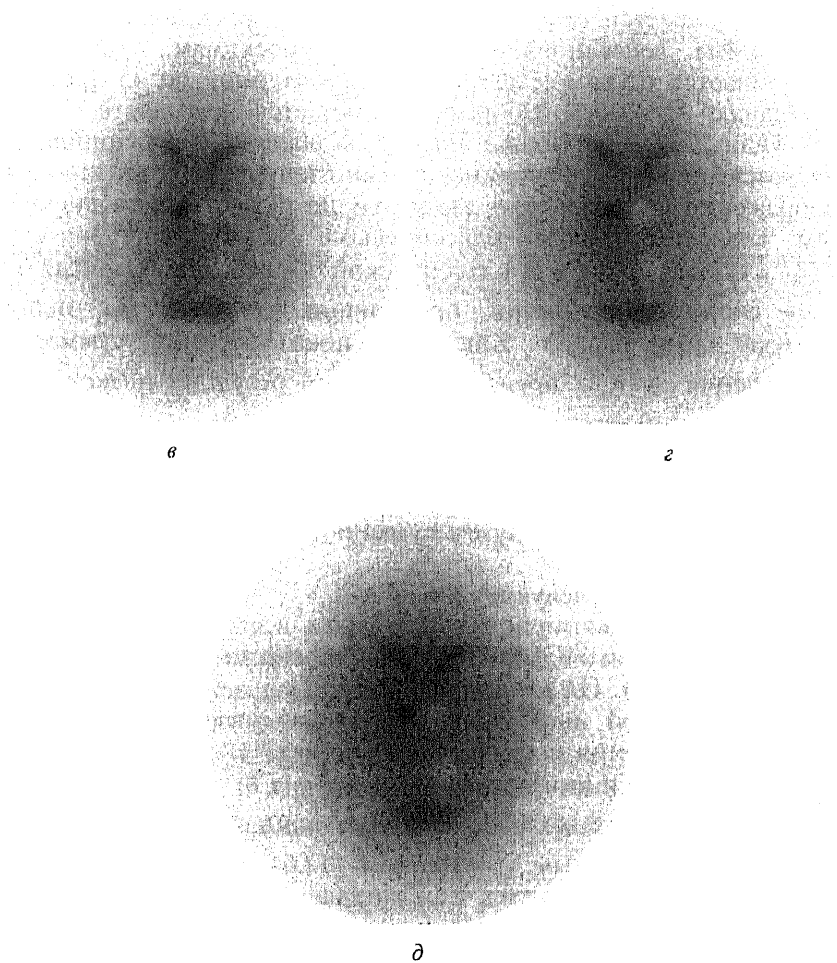


Рис. 12.9. Окончание

На рис. 12.9, в...д показаны результаты применения алгоритма обратного проецирования для восстановления изображения «Фантом» (рис. 12.9, а) по 15, 45 и 180 равноотстоящим по углу  $\theta$  проекциям. Изображения типа «Фантом» обычно используются для тестирования различных алгоритмов восстановления. Изображение радоновского образа для 180 проекций приведено на рис. 12.9, б. Число лучей равно 128. Полосы на восстановленных томограммах уже не заметны, и практически все детали

рис. 12.9, а можно рассмотреть. Качество восстановления значительно улучшается при увеличении числа проекций с 15 до 45. Томограмма, восстановленная по 180 проекциям, практически не отличается от томограммы, восстановленной по 45 проекциям. Однако четкость изображения остается неудовлетворительной. Использование только алгоритма обратного проецирования для восстановления, очевидно, должно приводить к размыванию изображения, так как этот алгоритм по сути является дискретным аналогом метода классической томографии. Значение  $u_{\Sigma}(x_k, y_l)$ , так же как и в классической томографии, вычисляется путем сложения проекций, проходящих через как бы «неподвижную» точку  $(x_k, y_l)$ . Как будет показано ниже, суммарное изображение  $u_{\Sigma}(x, y)$  представляет собой результат низкочастотной фильтрации исходного изображения  $u(x, y)$ .

Таким образом, идея алгоритма обратного проецирования состоит в том, что оценку плотности  $u(x, y)$  в любой точке с координатами  $(x, y)$  находят путем суммирования лучей, проходящих через эту точку.

Можно сделать следующие выводы.

1. Метод компьютерной томографии всегда состоит из двух этапов. На первом — формируются проекционные данные (радоновский образ). На втором — с помощью некоторого алгоритма по полученной информации восстанавливается изображение поперечного сечения исследуемого объекта.

2. Между проекционными данными  $p(s, \theta)$  и восстанавливаемым изображением  $u(x, y)$  существует однозначная связь, определяемая интегральным уравнением (12.6) или (12.8). Поэтому для нахождения алгоритма восстановления томограммы  $u(x, y)$  по  $p(s, \theta)$  необходимо найти решение интегрального уравнения. Изображения, полученные методом обратного проецирования, имеют низкую четкость, так как не являются этим решением.

3. Полученные проекционные данные представляют собой дискретное изображение  $p(s, \theta)$ , значения отсчетов которого равны интегралам (12.6) или (12.8). Дискретность изображения  $p(s, \theta)$  по параметрам  $s$  и  $\theta$  обусловлена технической реализацией томографов. Конечное число интегралов, очевидно, приведет к погрешностям восстановления томограммы.

## 12.4. ТЕОРЕМА О ЦЕНТРАЛЬНОМ СЕЧЕНИИ

В основе большинства алгоритмов восстановления томограмм лежит теорема о центральном сечении, которая устанавливает связь между преобразованием Фурье функций  $u(x, y)$  и  $p(s, \theta)$ . Далее будем обозначать: одномерное прямое и обратное преобразование Фурье символами  $F_1\{\cdot\}$  и  $F_1^{-1}\{\cdot\}$ , а двумерное —  $F_2\{\cdot\}$  и  $F_2^{-1}\{\cdot\}$ ; фурье-образы от функций  $u(x, y)$  и  $p(s, \theta)$  — строчными буквами, причем нижний индекс будет обозначать размерность преобразования, например:

$$P_1(f, \theta) = F_1\{p(s, \theta)\} = \int_{-\infty}^{\infty} p(s, \theta) \exp(-i2\pi s f) ds; \quad (12.17)$$

$$p(s, \theta) = F_1^{-1}\{P_1(f, \theta)\} = \int_{-\infty}^{\infty} P_1(f, \theta) \exp(i2\pi s f) df;$$

$$\begin{aligned} U_2(f_x, f_y) &= F_2\{u(x, y)\} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) \exp(-i2\pi(f_x x + f_y y)) dx dy; \end{aligned} \quad (12.18)$$

$$\begin{aligned} u(x, y) &= F_2^{-1}\{U_2(f_x, f_y)\} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_2(f_x, f_y) \exp(i2\pi(f_x x + f_y y)) df_x df_y, \end{aligned}$$

где  $f$ ,  $f_x$  и  $f_y$  — линейные частоты,  $i$  — мнимая единица.

Следует отметить, что в отличие от  $(s, \theta)$ , пара вещественных чисел  $(f, \theta)$  из области определения фурье-образа проекции Радона  $P_1(f, \theta)$  может быть интерпретирована как точка на плоскости в полярных координатах, так как  $P_1(0, \theta_1) = P_1(0, \theta_2)$  при  $\theta_1 \neq \theta_2$  (докажите это). Кроме того, функция  $P_1(f, \theta)$  удовлетворяет условию

$$P_1(f, \theta) = P_1(-f, \theta + \pi), \quad (12.19)$$

аналогичному (12.7) или (12.10). Соотношение (12.19) можно доказать, воспользовавшись свойством (12.7) радоновских обра-

зов. Подставив в (12.17) функцию  $p(-s, \theta + \pi)$  вместо  $p(s, \theta)$ , а затем, заменив переменную  $s$  на  $-s$ , получим:

$$\begin{aligned} P_1(f, \theta) &= \int_{-\infty}^{\infty} p(-s, \theta + \pi) \exp(-i2\pi s f) ds = \\ &= \int_{-\infty}^{\infty} p(s, \theta + \pi) \exp(i2\pi s f) ds = P_1(-f, \theta + \pi). \end{aligned}$$

Двумерное обратное преобразование Фурье в полярных координатах, связанных с декартовыми координатами в частотной области соотношениями

$$\begin{cases} f_x = f \cos \theta; \\ f_y = f \sin \theta \end{cases} \quad (12.20)$$

и

$$\begin{cases} f = \sqrt{f_x^2 + f_y^2}; \\ \theta = \arctg \frac{f_y}{f_x}, \end{cases} \quad (12.21)$$

имеет вид

$$\begin{aligned} u(x, y) &= F_2^{-1}\{U_2(f \cos \theta, f \sin \theta)\} = \\ &= \int_0^{2\pi} \int_0^{\infty} f U_2(f \cos \theta, f \sin \theta) \exp(i2\pi f(\cos \theta x + \sin \theta y)) df d\theta. \end{aligned} \quad (12.22)$$

Здесь учтено, что якобиан преобразования (12.20) равен  $f$ .

Теорема о центральном сечении утверждает, что одномерный фурье-образ проекции  $p(s, \theta)$  равен сечению двумерного фурье-образа функции  $u(x, y)$  вдоль прямой, проходящей через начало координат в частотной области под углом  $\theta$  к оси  $f_x$ , т.е.

$$P_1(f, \theta) = U_2(f \cos \theta, f \sin \theta) = U_2(f_x, f_y), \quad (12.23)$$

где координаты в прямоугольной  $(f_x, f_y)$  и полярной  $(f, \theta)$  системах координат связаны соотношениями (12.20), (12.21). Иными словами, одномерный фурье-образ  $p(s, \theta)$  является центральным сечением двумерного фурье-образа функции  $u(x, y)$ .



Для доказательства (12.23) воспользуемся соотношением (12.6). Одномерное преобразование Фурье от  $p(s, \theta)$  по параметру  $s$  равно

$$P_1(f, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(s \cos \theta - y' \sin \theta, s \sin \theta + y' \cos \theta) \exp(-i2\pi s f) ds dy'. \quad (12.24)$$

Выполнив замену переменных в (12.24) в соответствии с уравнением прямой (12.4) и формулами преобразования координат (12.3), получим

$$\begin{aligned} P_1(f, \theta) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) \exp[-i2\pi f(x \cos \theta + y \sin \theta)] dx dy = \\ &= U_2(f \cos \theta, f \sin \theta) = U_2(f_x, f_y). \end{aligned}$$

Из (12.23) следует, что двумерный фурье-образ функции  $u(x, y)$  в полярной системе координат можно вычислить, выполнив ее преобразование Радона (12.5), а затем, осуществив одномерное преобразование Фурье проекций по переменной  $s$ . Преобразование Радона функции  $u(x, y)$  можно получить, вычислив ее двумерное преобразование Фурье в полярной системе координат  $\{f, \theta\}$  и взяв обратное одномерное преобразование Фурье по переменной  $f$ .

## 12.5. ФУРЬЕ-АЛГОРИТМ ВОССТАНОВЛЕНИЯ ТОМОГРАММ

Невысокая эффективность алгоритма обратного проектирования объясняется тем, что он является эвристическим (полученным опытным путем). Для того чтобы точно восстановить функцию  $u(x, y)$  по проекциям  $p(s, \theta)$ , необходимо найти преобразование, обратное преобразованию Радона.

По сути, для определения неизвестной функции  $u(x, y)$  надо решить интегральное уравнение (12.6) или (12.8). Впервые такое решение было предложено Радонам. Одной из возможных реализаций этого решения является фурье-алгоритм, использующий теорему о центральном сечении. Найдем последовательность операций, реализующих этот алгоритм.

По двумерному фурье-образу  $U_2(f \cos \theta, f \sin \theta)$  можно найти саму функцию  $u(x, y)$ , используя обратное двумерное преобразо-

вание Фурье в полярных координатах (12.22). Подставив (12.23) в (12.22), получим

$$u(x, y) = \int_0^{2\pi} \int_0^{\infty} P_1(f, \theta) \exp(i2\pi f(x \cos \theta + y \sin \theta)) df d\theta. \quad (12.25)$$

Таким образом, в фурье-алгоритме восстановления томограмм вначале вычисляются одномерные фурье-образы  $P_1(f, \theta)$  по проекциям  $p(s, \theta)$  (см. 12.17). Получаем двумерный спектр томограммы в **полярной системе координат**  $(f, \theta)$ . Затем выполняем обратное двумерное преобразование Фурье также в **полярной системе координат**.

Основная трудность в использовании фурье-алгоритма возникает при его применении к реальным данным. Напомним, что проекционные данные имеют дискретный характер, так как вычисляются для дискретных значений  $s$ ,  $\theta$  и представляют собой массив значений проекций  $p(s_i, \theta_j)$ ,  $i = 0, N_s - 1$ ;  $j = 0, N_\theta - 1$ . При реализации фурье-алгоритма по дискретным данным используются процедуры одномерного и двумерного ДПФ или БПФ. Однако эти преобразования осуществляются на **прямоугольной сетке** отсчетов, а фурье-образы проекций  $p(s_i, \theta_j)$  мы получим на **полярной сетке** (рис. 12.10).

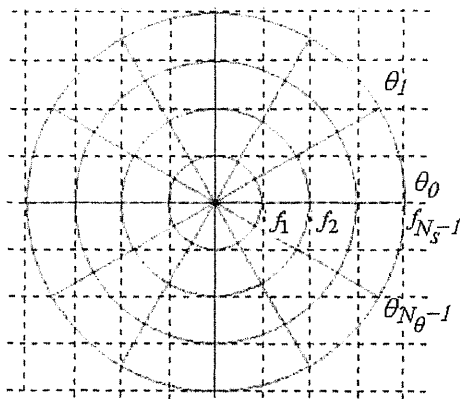


Рис. 12.10. Прямоугольная и полярная сетки отсчетов в пространстве частот

Поэтому для оценки значений  $P_1(f_{xk}, f_{yl})$  в узлах прямоугольной сетки (где  $k = 0, N_x - 1$ ;  $l = 0, N_y - 1$ ) по известным зна-

чениям  $P_1(f_i, \theta_j)$  в узлах полярной сетки необходима та или иная интерполяция. Здесь можно использовать как простейшие методы интерполяции (например, интерполяция по ближайшему значению), так и метод, реализуемый с помощью БПФ [12.6], который применяется для интерполяции в радиальном направлении.

При использовании БПФ каждую проекцию в пространственной области предварительно дополняют нулями. Затем находят фурье-образ таких проекций. Чем больше нулей добавлено в исходную  $N_s$  — точечную проекцию, тем меньше шаг дискретизации по частоте  $f$  (почему?) и, следовательно, тем больше число интерполированных значений  $P_1(s, f)$ . Интерполяцию с помощью БПФ иллюстрирует рис. 12.11.

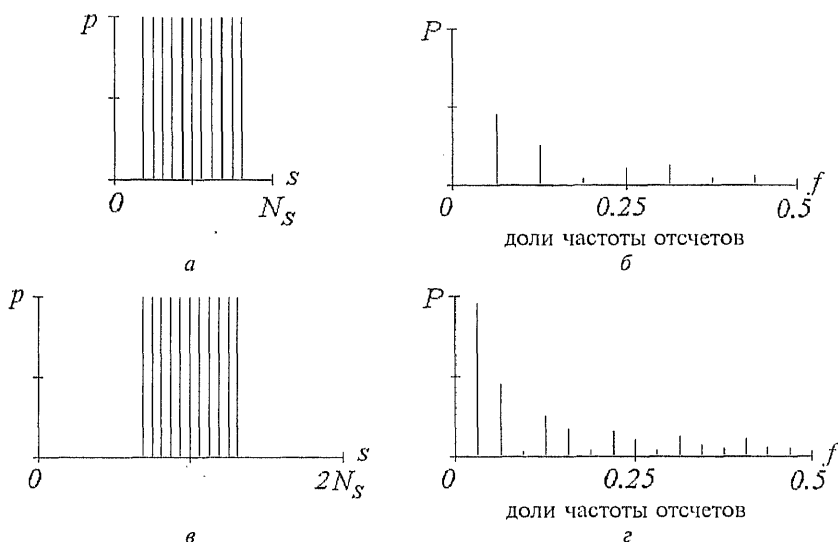


Рис. 12.11. Интерполяция за счет дополнения нулями:

*a* — исходная последовательность, содержащая шестнадцать отсчетов, одиннадцать из которых равны единице, остальные — нулю; *б* — модуль БПФ исходной последовательности; *в* — двукратно увеличенная исходная последовательность за счет дополнения нулями; *г* — модуль ее БПФ

Таким образом, фурье-алгоритм восстановления томограмм по дискретным проекционным данным состоит из трех основных операций:

1) вычисление одномерных дискретных фурье-образов проекций для дискретных значений угла  $\theta_j$ ;  $j = \overline{0, N_\theta - 1}$ ;

2) интерполяция значений  $P_1(f_{xk}, f_{yl})$  на прямоугольной сетке по значениям  $P_1(f_i, \theta_j)$  на полярной сетке;

3) дискретное или быстрое обратное преобразование Фурье на прямоугольной сетке, которое и дает оценку изображения  $u(x_k, y_l)$  также на прямоугольной сетке.

На рис. 12.12 показаны результаты восстановления изображения «Фантом» с помощью фурье-алгоритма. Видно, что в результате применения интерполяции устранено характерное для алгоритмов восстановления томограмм появление артефактов (ложных объектов и линий), которые особенно заметны на краях изображения. Использование 180 проекций и интерполяции с помощью БПФ (число отсчетов в исходной проекции было увеличено в четыре раза за счет дополнения нулями) позволило практически идеально восстановить изображение «Фантом».

Обсудим выбор параметров  $N_s$ ,  $N_\theta$ ,  $N_x$  и  $N_y$ . Очевидно, что величина шага дискретизации по пространственным переменным  $s$ ,  $x$  и  $y$  должна выбираться по теореме Котельникова исходя из величины верхней частоты  $f_v$  фурье-образов. Из теоремы о центральном сечении следует, что одномерный фурье-образ проекции определяет сечение («спицу», рис. 12.10) двумерного фурье-образа восстанавливаемой томограммы. Поэтому верхние частоты у одномерного фурье-образа проекции и двумерного фурье-образа томограммы равны.

Следовательно, шаги дискретизации  $\Delta s$  и  $\Delta$  по пространственным переменным  $s$  и  $(x, y)$  могут быть одинаковыми, при этом число отсчетов  $N = N_s = N_x = N_y = [2a/\Delta]$ . Если используется БПФ при реализации фурье-алгоритма, то дополнительным условием при выборе числа отсчетов  $N$  является его кратность степени 2. Увеличение шага дискретизации  $\Delta$  в пространственной области приводит к наложению спектров, а уменьшение  $\Delta$  (при этом число отсчетов  $N$  увеличивается) к возрастанию дозы облучения при получении дополнительных данных и объема вычислений, связанных с их обработкой.

При использовании БПФ количество отсчетов в частотной и пространственной области одинаково. Поэтому шаг по частоте  $\Delta f = \Delta f_x = \Delta f_y = \frac{2f_v}{N}$ . В соответствии с принципом дуальности, увеличение шага дискретизации  $\Delta f$  в частотной области приводит к наложению сигналов в пространственной области, при ко-

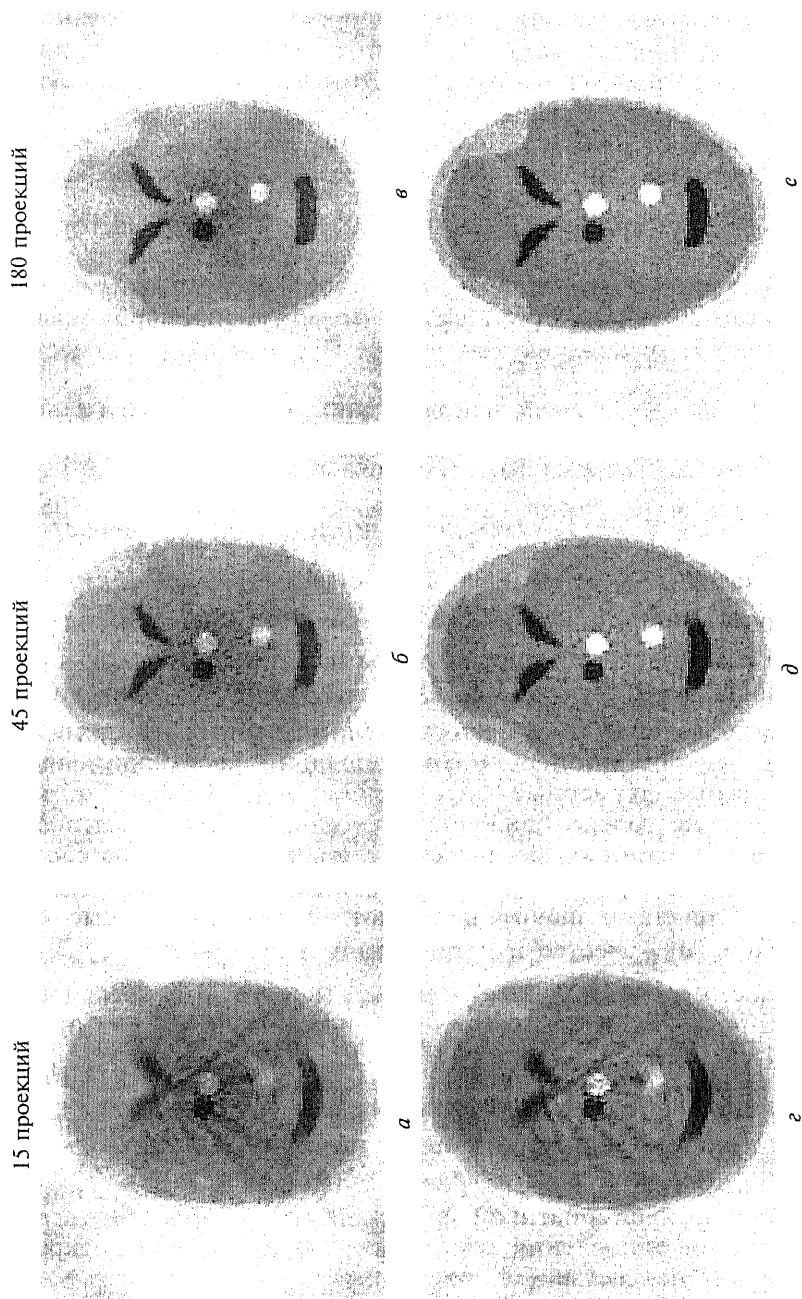


Рис. 12.12. Результат восстановления изображения фантомом фурье-алгоритмом:  
 а...е — восстановление с интерполяцией по ближайшему значению; г...е — восстановление с интерполяцией за счет  
 дополнения нулями (исходная последовательность была увеличена в 4 раза)

тором контрастные детали вблизи одного края восстанавливаемого изображения вызывают появление выбросов и осцилляций вблизи его противоположного края. Этот эффект особенно заметен на рис. 12.12, а...12.12, в.

При эквидистантном расположении проекций расстояние между «спицами» (см. рис. 12.10), на которых проекции задают

фурье-образ томограммы, равно  $\Delta f_0 = 2f \sin(\frac{\pi}{2N_0})$ , где частота  $f$

определяет радиус окружности, на которой расположены отсчеты. Следовательно, интервал между отсчетами пропорционален  $f$  и достигает максимума при  $f = f_b$ . При большом количестве

проекций  $\Delta f_0 \approx \frac{f\pi}{N_0}$ . Здесь использовано то, что  $\sin(\alpha) \approx \alpha$  при

малых углах  $\alpha$ . Теорема Котельникова требует, чтобы  $\Delta f_0 \leq \Delta f$ ,

поэтому  $N_0 \geq \frac{\pi}{2} N$ . Таким образом, число лучей  $N$  зависит от

верхней частоты  $f_b$  и размера  $a$  восстанавливаемой области, а число проекций  $N_0$  должно быть больше  $N$  примерно в полтора раза. На самом деле обеспечить такое большое количество проекций достаточно трудно, так как увеличение  $N_0$  приводит к увеличению времени сбора проекционных данных. Поэтому на практике вместо одномерной интерполяции часто применяют более сложные двумерные интерполяционные процедуры, использующие не только ближайшие соседние отсчеты, расположенные вдоль «спицы», но и отсчеты, расположенные на соседних «спицах». Из рис. 12.10 видно, что низкочастотная область заполнена отсчетами значительно плотнее, чем область высоких частот. Поэтому качество интерполяции значений  $P_1(f_{xk}, f_{yl})$  на низких частотах будет лучше, чем на высоких. В непрерывном случае этот недостаток компенсируется умножением  $P_1(f, \theta)$  на  $f$  (см. (12.25)). В фурье-алгоритме с использованием ДПФ или БПФ процедура умножения фурье-образов проекций на  $f$ , по сути, заменяется процедурой интерполяции. Это приводит к снижению четкости восстанавливаемого изображения  $u(x, y)$ . Следовательно, альтернативой фурье-алгоритму может быть алгоритм для восстановления томограмм, использующий дискретный аналог соотношения (12.25).

## 12.6. ВОССТАНОВЛЕНИЕ ТОМОГРАММ С ПОМОЩЬЮ ОБРАТНОГО ПРОЕКЦИРОВАНИЯ

Алгоритмы, основанные на методе обратного проектирования, нашли широкое применение в компьютерных томографах благодаря своей простоте и высокой точности. В их основе лежит соотношение (12.25).

Заменим область интегрирования в (12.25)

$$\begin{cases} 0 \leq \theta < 2\pi; \\ 0 \leq f < -\infty \end{cases}$$

на более удобную для реализации алгоритма восстановления

$$\begin{cases} 0 \leq \theta < \pi; \\ -\infty < f < -\infty, \end{cases} \quad (12.26)$$

используя свойство фурье-образа проекции Радона (12.19). В этом случае (12.25) можно представить в виде:

$$u(x, y) = \int_0^\pi \int_{-\infty}^\infty |f| P_1(f, \theta) \exp(i2\pi f(x \cos \theta + y \sin \theta)) df d\theta. \quad (12.27)$$

Заметим, что фурье-образ в полярных координатах двумерной функции  $u(x, y)$  в общем случае обладает свойством, подобным (12.19). Поэтому выражение (12.22) для обратного двумерного преобразования Фурье в полярных координатах можно также представить для области интегрирования (12.26) в виде

$$\begin{aligned} u(x, y) &= F_2^{-1}\{U_2(f \cos \theta, f \sin \theta)\} = \\ &= \int_0^\pi \int_{-\infty}^\infty |f| U_2(f \cos \theta, f \sin \theta) \exp(i2\pi f(\cos \theta x + \sin \theta y)) df d\theta. \end{aligned} \quad (12.28)$$

Таким образом, чтобы по проекциям  $p(s, \theta)$  восстановить функцию  $u(x, y)$ , необходимо в соответствии с (12.27) выполнить определенную последовательность операций:

1) вычислить фурье-образ  $P_1(s, \theta)$  проекции  $p(s, \theta)$  по следующей формуле:

$$P_1(f, \theta) = \int_{-\infty}^\infty p(s, \theta) \exp(-i2\pi s f) ds,$$

которая с учетом ограниченных размеров исследуемых объектов имеет вид:

$$P_1(f, \theta) = \int_{-a}^a p(s, \theta) \exp(-i2\pi s f) ds ;$$

2) умножить  $P_1(f, \theta)$  на  $|f|$ ;

3) найти модифицированные проекции

$$\tilde{p}(x \cos \theta + y \sin \theta, \theta) = \int_{-\infty}^{\infty} |f| P_1(f, \theta) \exp(i2\pi f (x \cos \theta + y \sin \theta)) df ,$$

вычислив обратное одномерное преобразование Фурье;

4) произвести интегрирование по углу  $\theta$ :

$$u(x, y) = \int_0^{\pi} \tilde{p}(x \cos \theta + y \sin \theta, \theta) d\theta . \quad (12.29)$$

Очевидно, что операция, описываемая соотношением (12.29), является операцией обратного проецирования.

Для дискретных данных модифицированные проекции вычисляют с помощью одномерного БПФ. Интегрирование в (12.29) заменяют операцией суммирования по  $\theta_j$ . Следует отметить, что при вычислении оценки томограммы на дискретной прямоугольной сетке в четвертом пункте также применяют процедуру интерполяции. Однако эта интерполяция осуществляется не в частотной, как в фурье-алгоритме, а в пространственной области.

### Сверточный алгоритм

Выполнение первых трех операций в предыдущем алгоритме для вычисления модифицированных проекций  $\tilde{p}(\cdot)$  можно заменить операцией свертки проекций  $p(s, \theta)$  (при фиксированном угле  $\theta$ ) с функцией  $h(s)$ :

$$\begin{aligned} \tilde{p}(x \cos \theta + y \sin \theta, \theta) &= \tilde{p}(s, \theta) = \\ &= \int_{-a}^a h(s_1) p(s - s_1, \theta) ds_1 = h(s) \otimes p(s, \theta), \end{aligned} \quad (12.30)$$

где  $h(s) = F_1^{-1}\{|f|\}$  — импульсная характеристика фильтра с частотной характеристикой  $|f|$ . Очевидно, что данный фильтр уси-



ливает верхние частоты. На рис. 12.13 показана импульсная характеристика фильтра для дискретных данных при  $|a|=64$ . В силу четности коэффициента передачи фильтра его импульсная характеристика также четная. Следовательно, фильтр, формирующий модифицированные проекции  $\tilde{p}(s, \theta)$ , является некаузальным. Число значащих отсчетов в импульсной характеристике  $h(i)$  дискретного фильтра невелико. Это свойство импульсной характеристики будет использовано ниже для восстановления фрагментов томограмм. Из (12.29) и (12.30) следует, что обратное преобразование Радона реализуется с помощью сверточного алгоритма в два этапа. На первом этапе выполняют свертку (12.30) проекций  $p(s, \theta)$  с импульсной характеристикой  $h(s)$  по переменной  $s$ . Результатом свертки являются модифицированные проекции  $\tilde{p}(s, \theta)$ . На втором этапе осуществляют обратное проецирование модифицированных проекций.

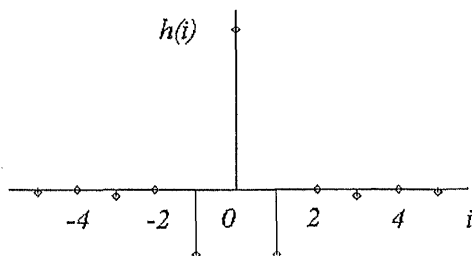


Рис. 12.13. Импульсная характеристика фильтра для вычисления модифицированных проекций

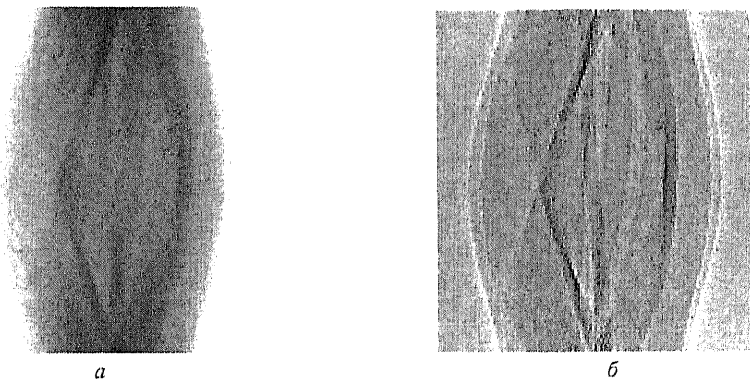
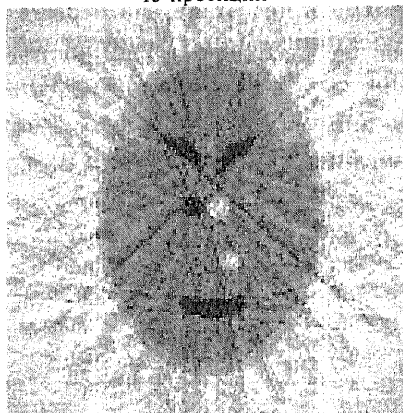


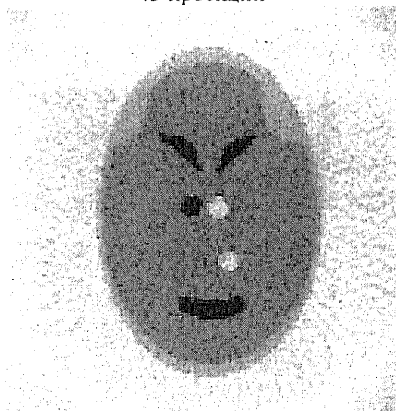
Рис. 12.14. Исходные и модифицированные проекции ( $N_0 = 180$ )

Исходные и модифицированные проекции изображения «Фантом» приведены на рис 12.14, *а* и *б*. Модифицированные проекции отличаются от исходных повышенной четкостью, обусловленной применением фильтра верхних частот  $|f|$ . На рис. 12.15, *а...в* показаны результаты восстановления изображения «Фантом» сверточным алгоритмом по 15, 45 и 180 проекциям. Неидеальность восстановления томограммы объясняется тем, что количество проекций, полученных под различными углами зондирования, и число лучей являются конечными.

15 проекций

*а*

45 проекций

*б*

180 проекций

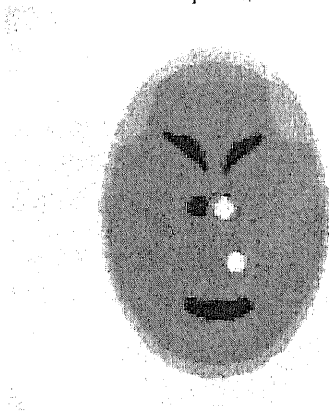
*в*

Рис. 12.15. Результат восстановления изображения «Фантом» сверточным алгоритмом

Для сравнения эффективности работы рассмотренных выше алгоритмов на рис. 12.16 приведены результаты восстановления реальной томограммы черепа с атрофией передней части мозга после тяжелой черепно-мозговой травмы (см. рис. 12.16, *а*). Это изображение далее будем называть «Томограмма». Особенностью изображения «Томограмма» является то, что в отличие от изображения «Фантом» у него флюктуирует яркость на однородных участках. Из приведенных данных следует, что при 180 проекциях фурье-алгоритм с интерполяцией и сверточный алгоритм дают вполне удовлетворительные результаты.

Рассмотренные алгоритмы получены на основе теоремы о центральном сечении для преобразования Фурье. При выводе фурье-алгоритма преобразование Фурье было записано в декартовой системе координат, а при выводе сверточного алгоритма — в полярной. Несмотря на общую основу, практическая реализация этих алгоритмов различна. Сверточный алгоритм оказался наиболее широко используемым алгоритмом в компьютерных томографах. Его преимуществом является то, что операции свертки и обратного проецирования для каждого ракурса могут выполняться независимо, а результирующее изображение представляет собой сумму изображений, полученных для каждого из ракурсов. Поэтому затраты машинного времени для ЭВМ с параллельной обработкой данных оказываются исключительно малыми. Анализ объемов вычислений, требуемых для реализации алгоритмов, показал, что фурье-алгоритм экономичнее алгоритма свертки за счет использования двумерного БПФ. Однако при его реализации необходимо массивы исходных данных дополнять нулями, чтобы уменьшить ошибки интерполяции. Следует отметить, что экономичная процедура БПФ может быть реализована для размеров строк и столбцов изображений, кратных  $2^k$ . Это ограничение приводит также к увеличению объема вычислений при реализации фурье-алгоритма.

Кроме того, сверточный алгоритм позволяет построить томограмму локального участка исследуемого объекта достаточно высокого качества. Задача восстановления локального участка возникает в тех случаях, когда неудобно или нежелательно собирать проекционные данные по всему сечению тела. Например, при исследовании биологических тканей человека целесообразно уменьшить дозу облучения, ограничив воздействие зондирующего пучка лучей той частью сечения, которая представляет интерес для медицинской диагностики. В этом случае просвечивается не вся область, ограниченная окружностью радиусом  $a$  (см. рис. 12.1),

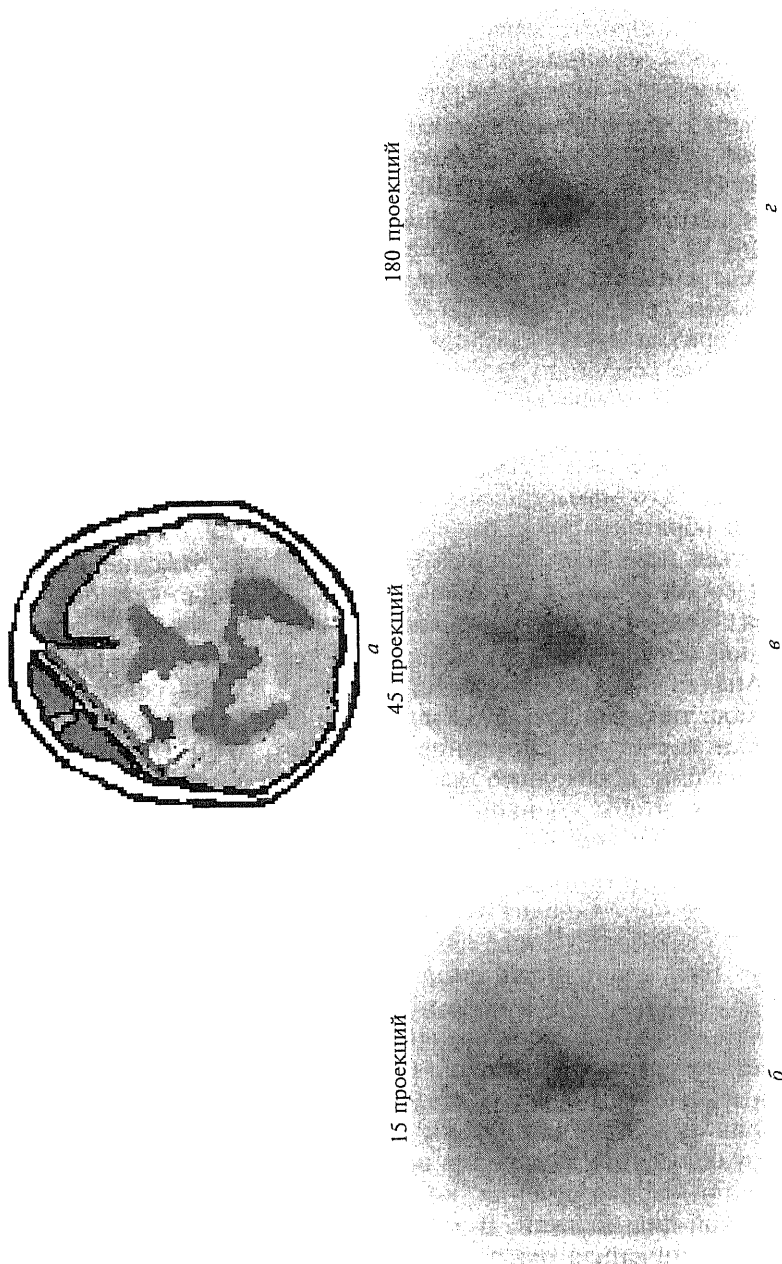


Рис. 12.16. Результаты восстановления изображения «Томограмма»:  
 а — исходное изображение «Томограмма»; б...г — алгоритм обратного преобразования; д...ж — фурье-алгоритм с интерполяцией по ближайшему значению; з...к — фурье-алгоритм с интерполяцией за счет дополнения нулями; исходная последовательность была увеличена в 4 раза; л...н — сверточный алгоритм (см. также с. 335, 336)

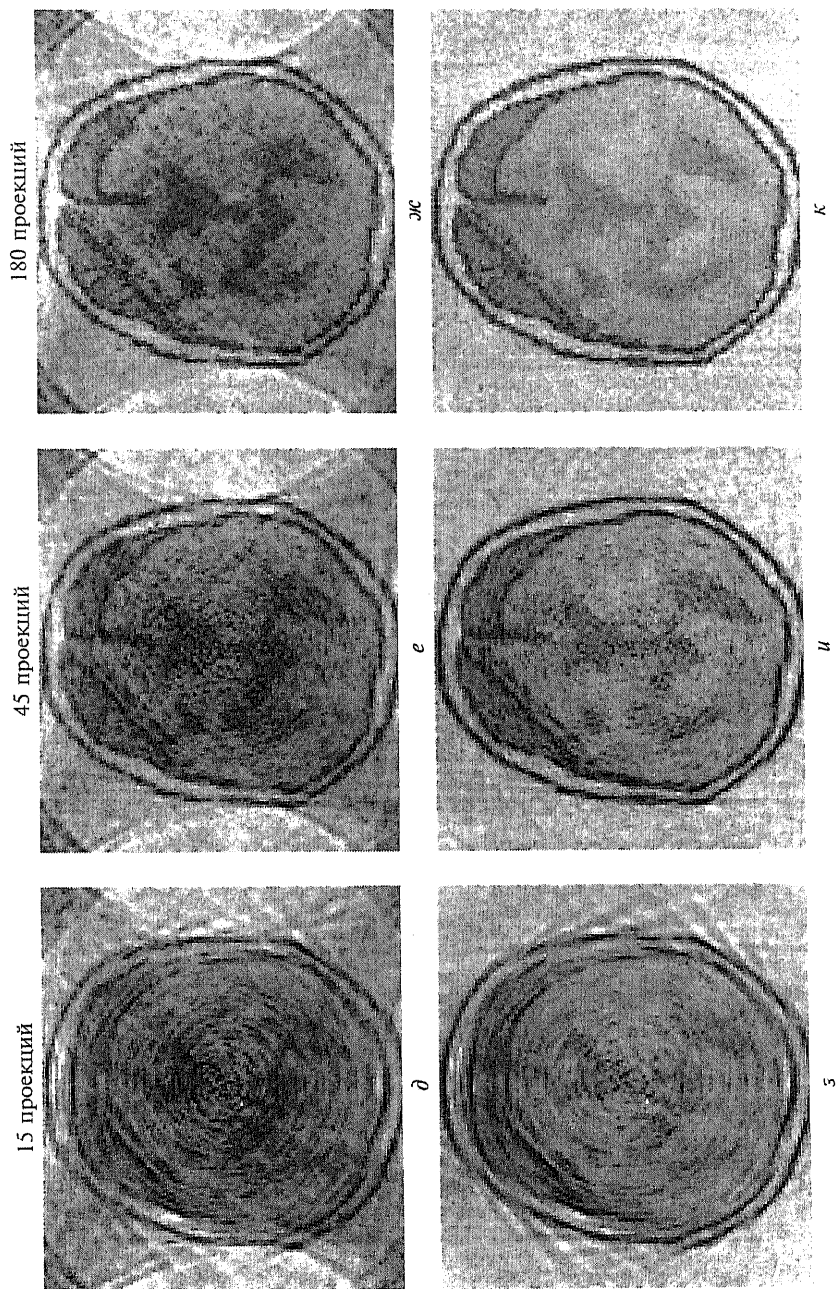


Рис. 12.16. Продолжение

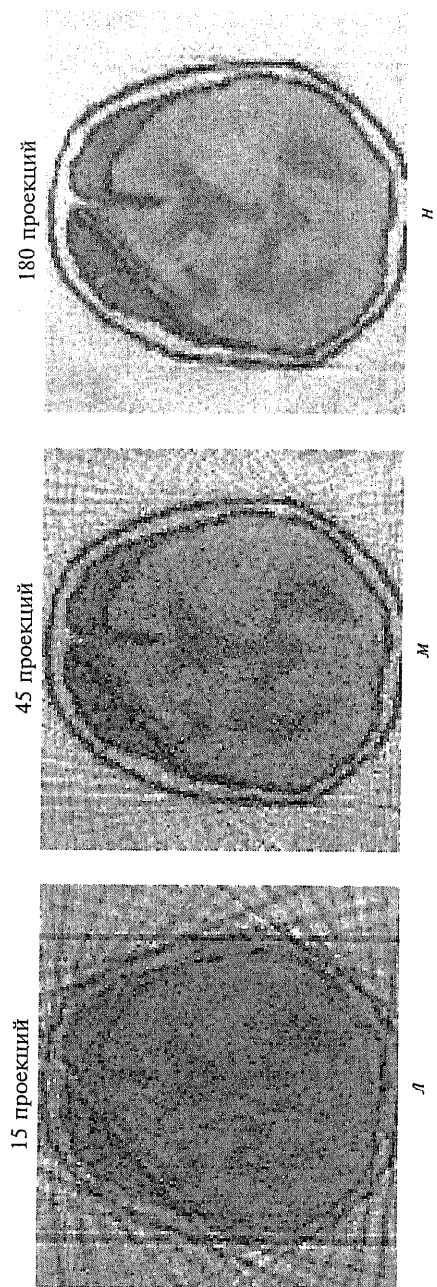


Рис. 12.16. Окончание

а лишь та ее часть, где находится интересующий исследователя локальный участок. При этом количество лучей, а следовательно, и уровень облучения могут быть значительно уменьшены.

На рис. 12.17 приведены результаты восстановления центральной части изображения «Фантом» сверточным и фурье-алгоритмом при 180 проекциях и 50 лучах, проходящих через центральную часть изображения. Радоновский образ для фиксированных значений  $s$  и  $\theta$  вычислялся путем суммирования данных вдоль всего луча, проходящего через изображение размером  $128 \times 128$  элементов. Поэтому отсчеты исходного изображения, которые находились за пределами локальной области, являлись по сути помехой. Восстанавливалось изображение размером  $50 \times 50$  элементов. Естественно, удастся получить изображение лишь локальной области, представляющей собой окружность, диаметр которой зависит от количества лучей. Сверточный алгоритм обеспечивает более высокое качество восстановления, чем фурье-алгоритм.

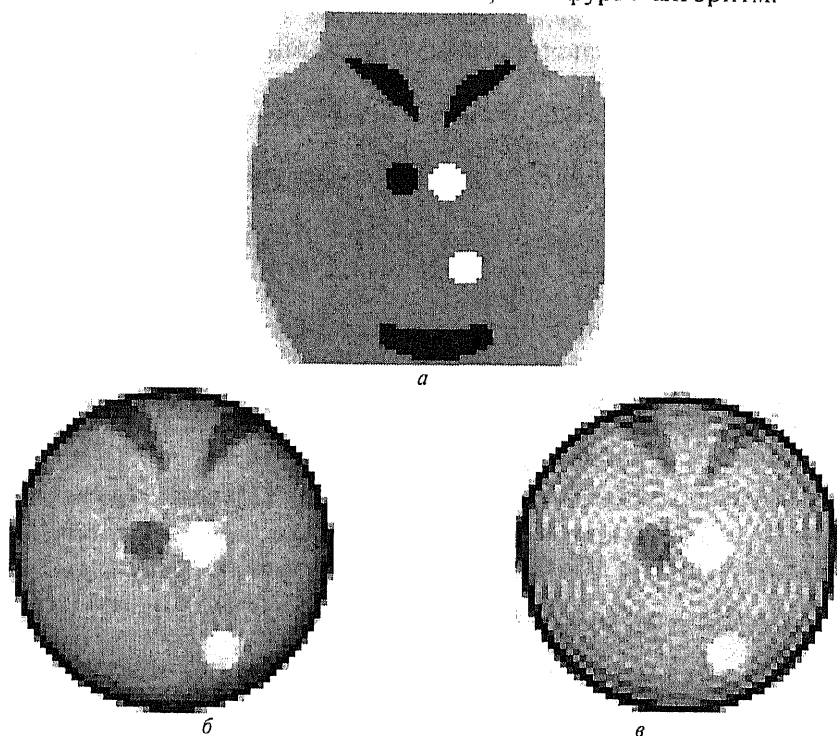


Рис. 12.17. Результаты восстановления локального фрагмента изображения фантом (размеры всех изображений увеличены в 2 раза):  
а – центральная часть исходного изображения «Фантом»; б и в – результаты восстановления локального фрагмента изображения «Фантом» сверточным и фурье алгоритмами

Это объясняется тем, что импульсная характеристика фильтра содержит лишь несколько значащих отсчетов. Поэтому отсутствие проекционных данных за пределами сравнительно узкого пучка приводит лишь к искажению краев локальной области. Количество искаженных отсчетов на краях локальной области определяется числом значащих отсчетов в импульсной характеристике фильтра. При восстановлении с помощью фурье-алгоритма на изображении наблюдаются достаточно сильные высокочастотные осцилляции яркости, обусловленные резким скачком яркости на краях проекционных данных. Для их устранения необходимо применять регуляризующие окна, аналогичные тем, что рассмотрены в четвертой главе.

### 12.7. ИТЕРАЦИОННЫЕ АЛГОРИТМЫ ВОССТАНОВЛЕНИЯ ТОМОГРАММ

Для решения задачи восстановления томограмм по проекциям могут быть использованы итерационные алгоритмы, рассмотренные в главе 4.

Напомним, что итерационные алгоритмы обладают следующими существенными достоинствами:

1) при их построении не требуется определять обратный оператор;

2) достаточно просто синтезируются нелинейные итерационные алгоритмы, учитывающие априорную информацию о восстанавливаемом изображении;

3) при реализации этих методов возможна работа в интерактивном режиме, позволяющая сделать компромиссный выбор между качеством восстановления и временем обработки.

Прежде чем рассмотреть итерационные методы восстановления томограмм, найдем связь между томограммой  $u(x, y)$  и суммарным изображением  $u_{\Sigma}(x, y)$ . Двумерное обратное преобразование Фурье в полярных координатах определяется соотношением (12.28), которое для удобства приведем здесь повторно

$$u(x, y) = F_2^{-1}\{U_2(f \cos \theta, f \sin \theta)\} = \int_0^{\pi} \int_{-\infty}^{\infty} |f| U_2(f \cos \theta, f \sin \theta) \exp(i2\pi f(\cos \theta x + \sin \theta y)) df d\theta. \quad (12.31)$$

С другой стороны, суммарное изображение характеризуется соотношением (12.22), которое с учетом (12.17) и (12.23) имеет вид



$$\begin{aligned}
 u_{\Sigma}(x, y) &= \int_0^{\pi} F_1^{-1}\{P_1(f, \theta)\} d\theta = \\
 &= \int_0^{\pi} \int_{-\infty}^{\infty} U_2(f \cos \theta, f \sin \theta) \exp(i2\pi f(x \cos \theta + y \sin \theta)) df d\theta.
 \end{aligned} \tag{12.32}$$

Сравнивая (12.31) и (12.32), получаем выражение, определяющее связь между двумерными фурье-образами суммарного изображения  $U_{\Sigma,2}(\cdot)$  и томограммой  $U_2(\cdot)$ :

$$U_{\Sigma,2}(f \cos \theta, f \sin \theta) = \frac{U_2(f \cos \theta, f \sin \theta)}{|f|},$$

которое, в декартовых координатах имеет вид

$$U_{\Sigma,2}(f_x, f_y) = U_2(f_x, f_y) H_{\Sigma}(f_x, f_y), \tag{12.33}$$

где  $H_{\Sigma}(f_x, f_y) = \frac{1}{\sqrt{f_x^2 + f_y^2}}.$

Из (12.33) следует, что суммарное изображение представляет собой результат низкочастотной фильтрации томограммы, так как фильтр  $H_{\Sigma}(f_x, f_y)$  ослабляет верхние частоты. Поэтому суммарное изображение получается расфокусированным (или нечетким, см. рис. 12.9 и 12.16). Коэффициент передачи инверсного фильтра для восстановления томограммы по суммарному изображению имеет вид

$$H_{\text{инв}}(f_x, f_y) = \frac{1}{H_{\Sigma}(f_x, f_y)} = \sqrt{f_x^2 + f_y^2}. \tag{12.34}$$

Таким образом, возможен еще один алгоритм для восстановления томограмм:

- 1) по проекциям вычисляем суммарное изображение  $u_{\Sigma}$  в соответствии с формулой (12.15);
- 2) выполняем инверсную фильтрацию в частотной или в пространственной области

$$u(x, y) = u_{\Sigma}(x, y) \otimes h_{\text{инв}}(x, y),$$

где  $h_{\text{инв}}(x, y) = F_2^{-1}\{H_{\text{инв}}(f_x, f_y)\}$  — импульсная характеристика инверсного фильтра.

Соотношение (12.34) справедливо для бесконечного числа проекций. На практике его можно использовать, если число проекций превышает несколько десятков. При небольшом числе проекций инверсный фильтр может быть найден на основании методики определения коэффициента передачи  $H_{\Sigma}(\cdot)$  для дискретных данных, которая рассмотрена, например, в работе [12.2]. На рис. 12.18 приведены результаты инверсной фильтрации суммарных изображений «Томограмма» (рис. 12.18, а) и «Фантом» (рис. 12.18, б).

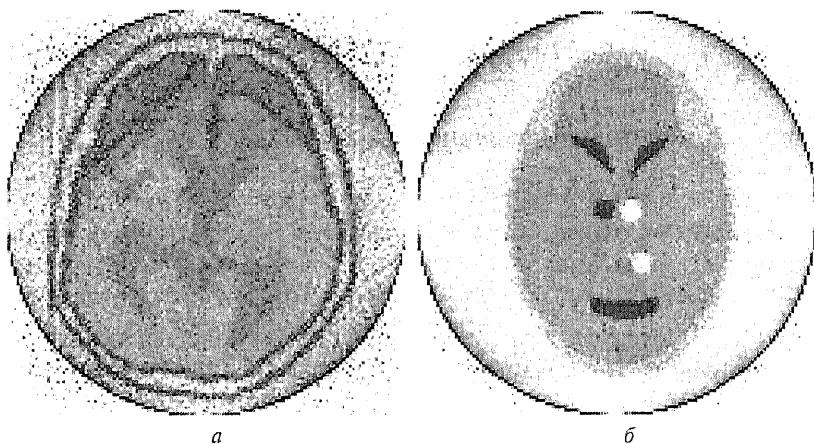


Рис. 12.18. Результаты восстановления томограмм инверсным фильтром по 45 проекциям

Поскольку суммарное изображение является расфокусированным изображением томограммы, для ее восстановления могут быть использованы итерационные алгоритмы, рассмотренные, например, в главе 4.

Линейный алгоритм Ван Циттерта для восстановления томограмм имеет вид

$$\hat{u}^{(0)}(x, y) = \alpha u_{\Sigma}(x, y); \quad (12.35)$$

$$\hat{u}^{(n)}(x, y) = \alpha u_{\Sigma}(x, y) + (\delta(x, y) - \alpha h_{\Sigma}(x, y)) \otimes \hat{u}^{(n-1)}(x, y),$$

где  $h_{\Sigma}(x, y) = F_2^{-1}\{H_{\Sigma}(f_x, f_y)\}$  — импульсная характеристика фильтра, формирующего суммарное изображение из томограммы;  $\alpha$  — коэффициент, определяющий сходимость итерационного алгоритма и влияющий на ее скорость.

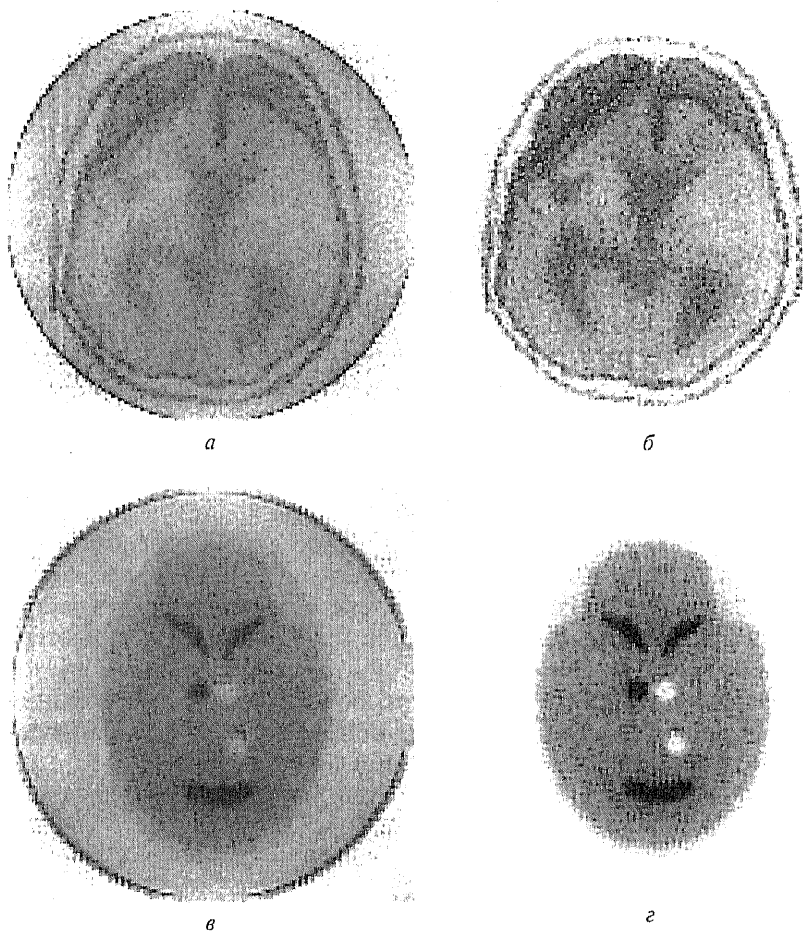


Рис. 12.19. Результаты восстановления томограмм линейным и нелинейным итерационными алгоритмами

Аналогичным образом задается нелинейный итерационный алгоритм, учитывающий априорные данные о восстанавливаемом изображении:

$$\begin{aligned}
 \hat{u}^{(0)}(x, y) &= \alpha u_{\Sigma}(x, y); \\
 \hat{u}^{(n)}(x, y) &= \alpha u_{\Sigma}(x, y) + \\
 &+ (\delta(x, y) - \alpha h_{\Sigma}(x, y)) \otimes \otimes \mathfrak{I}_1 \{ \mathfrak{I}_2 \{ \hat{u}^{(n-1)}(x, y) \} \},
 \end{aligned}
 \tag{12.35'}$$

где

$$\mathfrak{S}_1\{u(x, y)\} = \begin{cases} u(x, y); & \text{если } \sqrt{x^2 + y^2} \leq a, \\ 0; & \text{если } \sqrt{x^2 + y^2} > a, \end{cases}$$

$$\mathfrak{S}_2\{u(x, y)\} = \begin{cases} u(x, y); & \text{если } u(x, y) \geq 0, \\ 0; & \text{если } u(x, y) < 0 \end{cases}$$

— операторы пространственного ограничения и ограничения на неотрицательность томограммы.

На рис. 12.19 приведены результаты восстановления изображений «Томограмма» и «Фантом» линейным (рис. 12.19, *а* и *в*) и нелинейным (рис. 12.19, *б* и *г*) итерационными алгоритмами по 45 проекциям при  $\alpha = 0.15$  и для ста итераций. Очевидно, что применение нелинейного алгоритма позволяет повысить качество восстановления томограммы, особенно на краях.

Следует напомнить, что общим недостатком итерационных алгоритмов является их низкая вычислительная эффективность, обусловленная итеративным характером вычислений. Однако ряд итерационных алгоритмов нашел применение в компьютерной томографии. К сожалению, отсутствуют формальные подходы, позволяющие определить целесообразность использования именно итерационных алгоритмов. Вопросы о том, когда их следует использовать и сколько итераций необходимо выполнить, решаются в каждом конкретном случае исходя из практического опыта.

## 12.8. ФЛЮКТУАЦИОННЫЕ ИСКАЖЕНИЯ ПРОЕКЦИОННЫХ ДАННЫХ

Проекционные данные, регистрируемые в реальных компьютерных томографах, носят случайный характер, обусловленный следующими важными факторами. Во-первых, используемое рентгеновское излучение имеет ярко выраженную квантовую структуру. Число квантов, испускаемых рентгеновским источником за фиксированный интервал времени, является случайной величиной, поэтому элемент случайности заложен в самой природе источника излучения [12.1]. Во-вторых, проходя через исследуемый объект, кванты поглощаются веществом случайным образом [12.1]. Особенно случайный характер проекционных данных проявляется при низких уровнях облучения исследуемых объектов. В-третьих, всегда существуют внешний фон и внутренние

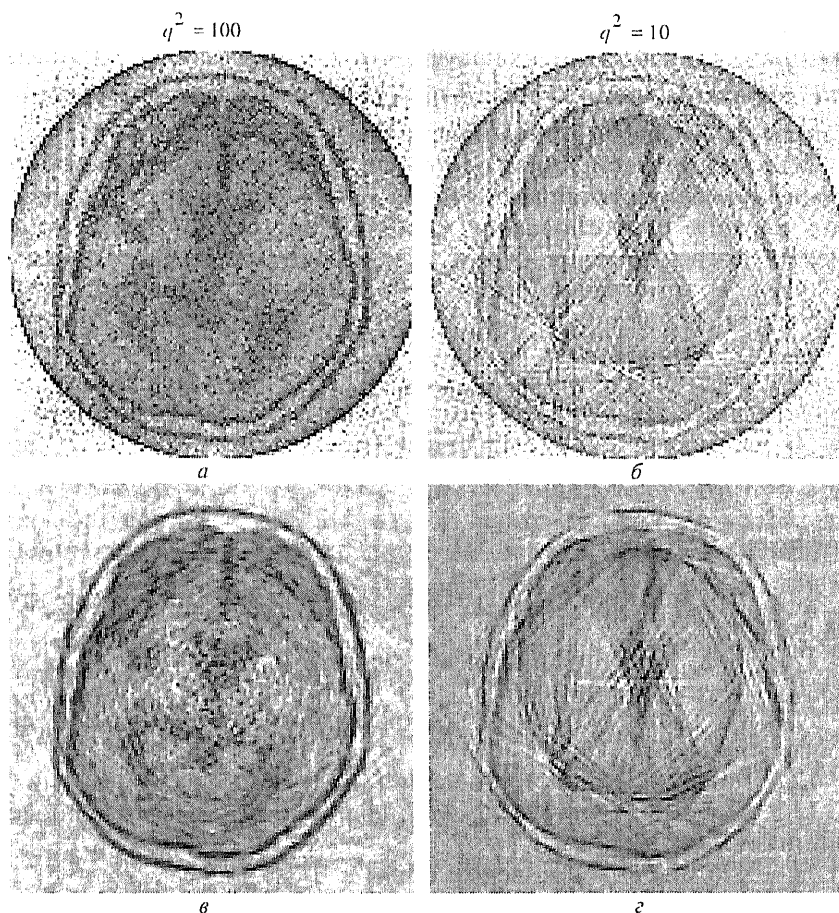


Рис. 12.20. Результаты восстановления томограмм при различных отношениях сигнал/шум

шумы регистрирующего устройства, например шумы квантования в компьютерных томографах. Рассмотренные выше алгоритмы восстановления томограмм по проекциям являются по сути различными вариантами реализации инверсной фильтрации, которая обладает низкой помехоустойчивостью. На рис. 12.20 показаны результаты восстановления изображения «Томограмма» методом инверсной фильтрации (рис. 12.20, а и 12.20, б) и фурье-алгоритмом (рис. 12.20, в и 12.20, г) при различных отношениях сигнал/шум  $q^2 = \frac{\sigma_p^2}{\sigma_\eta^2}$ , где  $\sigma_p^2$  и  $\sigma_\eta^2$  — дис-

персии проекций и аддитивного дельта-коррелированного шума. Из приведенных данных следует, что качественное восстановление возможно при  $q^2 > 100$ . Для борьбы с шумами используются методы фильтрации, описанные в главе 3.

Мы рассмотрели лишь основные алгоритмы восстановления томографических изображений. За время, прошедшее с появления первого рентгеновского компьютерного томографа, томография превратилась в бурно развивающуюся область науки и техники. В томографах стали использоваться новые источники излучения, появились новые принципы формирования томографических изображений и, естественно, новые математические методы восстановления изображений. Проблемы восстановления томографических изображений широко освещаются в литературе. В частности, ее математическим, техническим и вычислительным аспектам посвящены работы [12.1, 12.2, 12.5–12.8].

## ВОПРОСЫ К ГЛАВЕ 12

12.1. Что понимается под изображением внутренней структуры объекта?

12.2. Как формируется набор проекций для описания внутренней структуры объекта?

12.3. Из каких основных этапов состоит метод вычислительной томографии?

12.4. Как описываются измерения, получаемые в томографии?

12.5. Почему регистрируемая информация называется проекционными данными?

12.6. Что такое радоновский образ?

12.7. Почему система координат, в которой задан радоновский образ, не является полярной?

12.8. Что такое обратная проекция и суммарное изображение?

12.9. Какая теорема лежит в основе фурье-алгоритма восстановления томограмм?

12.10. Почему суммарное изображение имеет низкую четкость?

12.11. Как из суммарного изображения получить томограмму?

12.12. Почему регистрируемая информация является дискретной и как выбираются интервалы дискретизации?

12.13. Какие искажения проявляются в томограмме вследствие дискретизации проекций?

12.14. Какие свойства томограммы используются в нелинейном итерационном алгоритме?

12.15. Почему проекционные данные в реальных томографах носят случайный характер?

## ЗАКЛЮЧЕНИЕ

Цифровая обработка изображений принадлежит к числу интенсивно развивающихся направлений научных исследований и прикладных разработок. Характерно многообразие прикладных областей и постоянно возникающих в них новых задач. Далеко не всегда в арсенале готовых разработок удастся найти методы и алгоритмы, при помощи которых могут быть решены эти новые задачи. Поэтому очень часто создание алгоритма осуществляется на уровне научного исследования. Неудивительно, что при этом нередко случаи использования и эвристических подходов, лишенных серьезной математической основы. Следует отметить, что среди них имеются блестящие решения, ряд из них нашел отражение в настоящем пособии. Вместе с тем потребности реальных разработок обычно предъявляют к алгоритмическому обеспечению требования максимального качества обработки. Достигнуть этого можно на основе применения оптимальных методов, опирающихся на математические модели изображений и теорию оптимальных решений. Поэтому основное внимание в пособии уделялось именно таким подходам и методам.

Авторы стремились сосредоточить внимание на тех задачах и методах их решения, которые являются актуальными для новых прикладных разработок и которые вместе с тем опираются на современные достижения в теории сигналов (в том числе многомерных) и их обработки, в теории распознавания образов. Авторы считают, что большие перспективы в данной области принадлежат вероятностно-статистическим методам как применительно к системам, работающим с оператором, так и для систем с автоматической обработкой визуальной информации. Это нашло отражение в содержании пособия.

Учебное пособие не претендует на полноту освещения проблемы цифровой обработки изображений. Это не только невозможно, в этом и нет необходимости. Изучение дисциплины, овладение ее методологией достигается при знакомстве с ключевыми разделами.

## ЛИТЕРАТУРА

### К главе 1

- 1.1. *Даджион Д., Мерсеро Р.* Цифровая обработка многомерных сигналов. — М.: Мир, 1988.
- 1.2. *Ярославский Л.П.* Введение в цифровую обработку изображений. — М.: Сов. радио, 1979.
- 1.3. *Прэтт У.* Цифровая обработка изображений. Кн.1. — М.: Мир, 1982.
- 1.4. *Методы передачи изображений. Сокращение избыточности / У.К. Прэтт, Д.Д. Сакрисон, Х.Г.Д. Мусманн и др.; Под ред. У.К.Прэтта.* — М.: Радио и связь, 1983.
- 1.5. *Джэйн А.К.* Сжатие видеoinформации // ТИИЭР. — Т. 69. — № 3.

### К главе 2

- 2.1. *Горбунов Б.А., Дементьев В.Н., Пяткин В.П.* Распознавание изображений в дистанционном зондировании // Автоматизированная обработка изображений природных комплексов Сибири. — Новосибирск: Наука, 1988.
- 2.2. *Левин Б.Р.* Теоретические основы статистической радиотехники. Кн.1. — М.: Сов. радио, 1966.
- 2.3. *Ярославский Л.П.* Введение в цифровую обработку изображений. — М.: Сов. радио, 1979.
- 2.4. *Прэтт У.* Цифровая обработка изображений. Кн.2. — М.: Мир, 1982.

### К главе 3

- 3.1. *Спектор А.А.* Рекуррентная фильтрация гауссовских дискретных полей, наблюдаемых в гауссовских шумах // Радиотехника и электроника. — 1994. — № 7.
- 3.2. *Даджион Д., Мерсеро Р.* Цифровая обработка многомерных сигналов. — М.: Мир, 1988.
- 3.3. *Левин Б.Р.* Теоретические основы статистической радиотехники. Кн.2. — М.: Сов. радио, 1975.
- 3.4. *Woods J.W.* Two-Dimensional Digital Signal Processing 1. Perlin e.a., 1981.
- 3.5. *Васильев К.К., Крашенинников В.Р.* Методы фильтрации многомерных случайных полей/ Саратов. гос. ун-т, — Саратов, 1990.



3.6. Грузман И.С., Микерин В.И., Спектор А.А. Двухэтапная фильтрация изображений на основе использования ограниченных данных // Радиотехника и электроника – 1995 – № 5.

3.7. Грузман И.С., Спектор А.А. Применение свойства условной независимости для симметричного сглаживания марковских процессов // Радиотехника и электроника. – 1997. – № 5.

3.8. Грузман И.С. Двухэтапная фильтрация бинарных изображений // Автометрия. – 1999. – № 3.

3.9. Быстрые алгоритмы в цифровой обработке изображений / Под ред. Т.С. Хуанга. – М.: Радио и связь, 1984.

#### К главе 4

4.1. Борн М., Вольф Э. Основы оптики. – М.: Наука, 1970.

4.2. Прэтт У. Цифровая обработка изображений. Кн.1. – М.: Мир, 1982.

4.3. Беллман Р. Введение в теорию матриц / Пер. с англ.; Под ред. В.Б. Лидского. – М.: Наука, 1976.

4.4. Реконструкция изображений: Пер. с англ. / Под ред. Г.Старка. – М.: Мир, 1992.

4.5. Прэтт У. Цифровая обработка изображений. Кн. 2. – М.: Мир, 1982.

4.6. Василенко Г.И., Тараторин А.М. Восстановление изображений. – М.: Радио и связь, 1986.

4.7. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. – М.: Мир, 1978.

4.8. Бейтс Р., Мак-Донелл М. Восстановление и реконструкция изображений. – М.: Мир, 1989.

4.9. Бендат Дж., Пирсол А. Прикладной анализ случайных данных: Пер. с англ. – М.: Мир, 1989.

4.10. Грузман И.С. Компенсация краевых эффектов при коррекции линейных искажений изображений // Автометрия. – 1995. – № 2.

4.11. Бьемон Ж., Лагендейк Л., Мерсеро Р.М. Итерационные методы улучшения изображений // ТИИЭР. – 1990. – № 5.

#### К главе 5

5.1. Хорн Б.К.П. Зрение роботов. – М.: Мир, 1989.

5.2. Прэтт У. Цифровая обработка изображений. – М.: Мир, 1982.

5.3. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. – М.: Наука, 1988.

5.4. Тышкевич Р.И., Феденко А.С. Линейная алгебра и аналитическая геометрия. – Минск: Высшая школа, 1968.

5.5. Абрамов В.В., Киричук В.С. и др. Реконструкция трехмерных поверхностей по двум проекциям при отслеживании камерой заданной точки сцены // Автометрия, 1998. – № 5. – С. 3.

5.6. Киричук В.С., Перетягин Г.И. Об установлении сходства фрагментов с эталоном // Автометрия. — 1986. — № 4. — С. 83.

5.7. Битюцкий О.И., Перетягин Г.И. Поиск и локализация реперных фрагментов при совмещении повторных снимков // Автометрия. — 1988. — № 3.

5.8. *Jahne B.* Digital Image Processing: Concepts, Algorithms, and Scientific Applications. — Perlin: Springer-Verlag, 1993.

5.9. *Mostafavi H., Smith F.W.* Image correlation with geometric distortion. — IEEE Transactions. — 1978. — V.AES-14.

5.10. *Keys R.G.* Cubic convolution interpolation for Digital Image Processing. — IEEE Trans. Acoust., Speech, Signal Processing, 1981. — Vol. ASSP-29. — P. 1153–1160.

5.11. *Parker A.J., Kenyon R., Troxel D.E.* Comparison of interpolating methods for Image Resampling. — IEEE Trans. Medical Imaging, 1983. — Vol. MI-2, No 1.

5.12. *Birchfield S.* An Introduction to Projective Geometry (for computer vision). 1998. <http://www.dai.ed.ac.uk/Cvonline/geom.htm>.

## К главе 6

6.1. Корн Г., Корн Т. Справочник по математике. — М.: Наука, 1973.

6.2. Бронштейн Н.Н., Семендяев К.А. Справочник по математике. — М.: Наука, 1966.

6.3. *Faugeras O.* Tree Dimensional Computer Vision. — MIT Press, 1993.

6.4. *Рao С.Р.* Линейные статистические методы и их применения. — М.: Наука, 1968.

6.5. *Бард Й.* Нелинейное оценивание параметров. — М.: Статистика, 1979.

6.6. *Hu X., Ahuja N.* Motion and Structure Estimation Using Long Sequence motion models. // Image and Vision Computing. — 1993. — V. 11. — № 9.

6.7. *Maybank S., Faugeras O. A.* Theory of Self-calibration of a Moving Camera. // International Journal of Computer Vision. — 1992. — V.8. — № 2.

6.8. *Хорн Б.К.П.* Зрение роботов. — М.: Мир, 1989.

6.9. *Popov S.A., Kirichuk V.S.* Algorithm of Estimation of the Geometric Parameters of the System of Two Projection Cameras by the Method of the Least Squares (MLS). // Pattern Recognition and Image Analysis. — 1999. — N 2. — P. 304.

6.10. *Прэтт У.* Цифровая обработка изображений. — М.: Мир, 1982.

6.11. *Март Д.* Зрение: информационный подход к изучению представления и обработки зрительных образов. — М.: Радио и связь, 1987.

## К главе 7

7.1. Бакут П.А., Колмогоров Г.С., Ворновицкий И.Э. Сегментация изображений: методы пороговой обработки // Зарубежная радиоэлектроника. — 1987. — № 10. — С. 6...24.

7.2. Прэтт У.К., Фожра О.Д., Гагалович А. Применение моделей стохастических текстур для обработки изображений // ТИИЭР — 1981. — Т. 69. — № 5.

7.3. Грузман И.С. Выделение однородных областей на изображениях // Статистические методы обработки изображений. — Новосибирск: Новосиб. электротехн. ин-т, 1993.

7.4. Гимельфарб Г.Л., Залесный А.В. Модели марковских случайных полей в задачах генерации и сегментации текстурных изображений/ Средства интеллектуализации кибернетических систем. — Киев, 1989.

7.5. Geman S., Geman D. Stochastic Relaxation, Gibbs Distributions and the Payesian Restoration of Images //IEEE Trans. Pattern Anal. Machine Intel., Pami-6, # 6. — P. 721-741, 1984.

## К главе 8

8.1.Бакут П.А., Колмогоров Г.С. Сегментация изображений: методы выделения границ областей.// Зарубежная радиоэлектроника. — 1987. — № 10. С. 25...45.

8.2. Виттих В.А., Сергеев В.В., Сойфер В.А. Обработка изображений в автоматизированных системах научных исследований. — М.: Наука, 1982.

8.3. Woods J.W. Two-dimensional discrete Marcovian fields. //IEEE Trans. Inform.Theory, 1972. Vol. 18.

8.4. Левин Б.Р. Теоретические основы статистической радиотехники. Кн.3. — М.: Сов. радио, 1976.

## К главе 9

9.1. Гуревич И.Б. Проблема распознавания изображений. Распознавание. Классификация. Прогноз. Математические методы и их применение: Ежегодник. Вып.1. — М.: Наука, 1989.

9.2. Белоцерковский О.М., Глазунов А.С., Щенников В.В. Компьютерное распознавание человеческих лиц// Зарубежная радиоэлектроника. — 1997. — no.8.

9.3. Rowley H.A., Baluja S., Kanade T. Neural-Pased Face Detection //IEEE Transactions on Pattern Analysis and Machine Intelligence. — Vol.20. — no.1. — 1998.

9.4. Дистанционное зондирование: количественный подход /Ш.М. Дэйвис, Д.А. Ландгребе, Т.Л. Филлипс и др. — М.: Наука, 1983.

9.5. Коуэлмен Г.Б., Эндрюс Х.С. Сегментация изображений при помощи автоматической классификации// ТИИЭР. -1979. - Т.67. - № 5.

9.6. Прэтт У. Цифровая обработка изображений. - М.: Мир, 1982.

9.7. Ту Дж., Гонсалес Р. Принципы распознавания образов. - М.: Мир, 1978.

9.8. Фукунага К. Введение в статистическую теорию распознавания образов. - М.: Наука, 1979.

9.9. Горелик А.Л., Скрипкин В.А. Методы распознавания. - М.: Высшая школа, 1989.

9.10. Шишов С.А. Классификаторы на основе нейронных структур// Зарубежная радиоэлектроника. - 1992. - № 8.

9.11. Горелик А.Л., Гуревич И.Б., Скрипкин В.А. Современное состояние проблемы распознавания. Некоторые аспекты. - М.: Радио и связь, 1985.

9.12. Фризен Б.Р. Оценки, энтропия, правдоподобие// ТИИЭР. - 1985. - Т. 73. - № 12.

9.13. Аренс Х., Лейтер Ю. Многомерный дисперсионный анализ. - М.: Финансы и статистика, 1985.

9.14. Деев А.Д. Асимптотические разложения распределения статистик дискриминантного анализа// Статистические методы классификации. - М.:1972. - Вып. 31.

9.15. *Schneiderman H.* A Statistical approach to 3D Object Detection applied to Faces and Cars//Ph.D.Thesis. Robotics Institute Carnegie Mellon University, Pittsburg. - May, 2000.

9.16. *Lantermann A.D., Miller M.I., Snyder D.L.* Automatic Target Recognition via the simulation of Infrared Scenes// Center for Imaging Science, Electronic Systems and Signal Research laboratory, Washington University, St.Louis, Missouri. <http://cis.wustl.edu>.

## К главе 10

10.1. Матерон Ж. Случайные множества и интегральная геометрия. - М.: Мир, 1978.

10.2. *Serra J.* Image Analysis and Mathematical Morphology. - London-New York: Academic Press, 1982.

10.3. *Heijmans H.* Mathematical Morphology: Basic Principles.//Proceedings of Summer School on Morphological Image and Signal Processing. - Zakopane, Poland, 1995 ([ftp://ftp.cwi.nl/pub/morphology/report/Heijmans\\_Zakopane\\_intro.ps.Z](ftp://ftp.cwi.nl/pub/morphology/report/Heijmans_Zakopane_intro.ps.Z)).

10.4. *Soille P.* Morphological Image Analysis. -Perlin, Heidelberg, New York: Springer-Verlag, 1999.

## К главе 11

11.1. Колмогоров А.Н. Теория передачи информации// Сессия АН СССР по научным проблемам автоматизации производства (15-20 октября 1956 г.) Пленарные заседания АН СССР. — М., 1957.

11.2. Ватолин Д.С. Алгоритмы сжатия изображений. Методическое пособие. Изд. отдел фак-та ВМиК МГУ им. М.В.Ломоносова. — 1999.

11.3. Климов А.С. Форматы графических файлов. С.-Петербург. ДиаСофт.,-1995.

11.4. Sonka M., Hlavac V., Boyle R. Image Processing, Analysis and Machine Vision. — <http://www.icaen.uiowa.edu/~dip/LECTURE/DataCompression.html>.

11.5. Прэтт У. Цифровая обработка изображений. М.: Мир. — 1982.

11.6. Mallat S.G. A theory of multiresolution signal decomposition: the wavelet representation// IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1989. — Vol.11. — P. 674-693.

11.7. Barnsley M., Hurd L. Fractal image compression. A. K. Peters Ltd., Weelesley M., 1993.

11.8. Wallace G.K. The JPEG still picture compression standard//Communication of the ACM. — 1991. — Vol. 34, no.4. — P. 30-40.

## К главе 12

12.1. Хермен Г. Восстановление изображений по проекциям. — М.: Мир, 1983.

12.2. Левин Г.Г., Вишняков Г.Н. Оптическая томография. — М.: Радио и связь, 1981.

12.3. Грузман И.С., В.С. Киричук, Косых В.П., Перетягин Г.И., Спектор А.А. Цифровая обработка изображений в информационных системах: Учебное пособие. — Новосибирск: Изд-во НГТУ, 2000.

12.4. Бронштейн И.Н., Семендяев К.А. Справочник по математике для инженеров и учащихся втузов — М.: Наука, 1986.

12.5. Троицкий И.Н. Статистическая теория томографии. М.: Радио и связь, 1982.

12.6. Марпл. — мл. С.Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990.

12.7. Пикалов В.В., Преображенский Н.Г. Реконструктивная томография в газодинамике и физике плазмы. — Новосибирск: Наука, 1987.

12.8. Реконструктивная и вычислительная томография // ТИИЭР (тематический выпуск). — 1983. — № 3. — Том 71.

**И.С. Грузман, В.С. Киричук, В.П. Косых,  
Г.И. Перетягин, А.А.Спектор**

**ЦИФРОВАЯ ОБРАБОТКА ИЗОБРАЖЕНИЙ  
В ИНФОРМАЦИОННЫХ СИСТЕМАХ**

**Учебник**

Редактор *И.Л. Кескевич*  
Технический редактор *Г.Е. Телятникова*  
Корректор *Л.Н. Ветчакова*  
Компьютерная верстка *Н.Ю. Адыгезалова*

---

Лицензия № ИД 04303 от 20.03.01. Подписано в печать 07.12.2001 г.  
Формат 60×90 1/16. Бумага офсетная. Тираж 2000 экз.  
Уч. - изд. л. 21,1. Печ. л. 22,0. Изд. № 119. Заказ № 856

---

Отпечатано в ГУП РПО СО РАСХН  
пос. Краснообск Новосибирской обл.