



ФЕДЕРАЛЬНАЯ ЦЕЛЕВАЯ ПРОГРАММА
«ГОСУДАРСТВЕННАЯ ПОДДЕРЖКА ИНТЕГРАЦИИ ВЫСШЕГО ОБРАЗОВАНИЯ
И ФУНДАМЕНТАЛЬНОЙ НАУКИ НА 1997 – 2000 ГОДЫ»

В.Б. Кашкин, А.И. Сухинин

ДИСТАНЦИОННОЕ ЗОНДИРОВАНИЕ ЗЕМЛИ ИЗ КОСМОСА

Цифровая обработка изображений

УДК 681.51:504(07)

ББК 26.82

К31

Рецензенты:

И.И. Варфоломеев, Л.М. Ускова, А.П. Шевырьков

Кашкин В.Б., Сухинин А.И.

К31 Дистанционное зондирование Земли из космоса. Цифровая обработка изображений: Учебное пособие. – М.: Логос, 2001. – 264 с.: ил.

ISBN 5-94010-138-0

Описываются принципы и технологии космического дистанционного зондирования Земли в оптическом и радиодиапазоне, представления изображений в цифровой форме, ввода изображений в ПЭВМ, визуализации и обработки цифровых изображений, а также методы распознавания образов. Приводятся сведения о географических информационных системах. В качестве примеров использования космических средств обсуждаются результаты дистанционного зондирования стихийных бедствий (лесных пожаров, затопов льда на реках, наводнений, штормовых ветров), а также состояние растительности, водоемов, снегового и ледового покрова. Приводятся методы исследования вертикальных профилей атмосферы, изучения облачности. Рассмотрена проблема атмосферного озона.

Для студентов высших учебных заведений, обучающихся по направлениям и специальностям: «Дистанционное зондирование Земли», «Геоинформационные системы», «Экология», «Цифровая обработка информации». Представляет интерес для ученых и специалистов, занимающихся дистанционным зондированием Земли и других объектов, а также сотрудников МЧС.

ББК 26.82

Издание осуществлено при финансовой поддержке Федеральной целевой программы «Государственная поддержка интеграции высшего образования и фундаментальной науки на 1997–2000 годы»

ISBN 5-94010-138-0

© Центр «Интеграция», 2001

ОГЛАВЛЕНИЕ

Предисловие	5
1. Физические средства контроля окружающей среды	7
1.1. Общая характеристика дистанционных методов	7
1.2. Физические основы дистанционного зондирования	9
1.2.1. Излучение энергии нагретыми телами	9
1.2.2. Отражение и рассеяние солнечной энергии поверхностью	12
1.2.3. Влияние атмосферы	15
1.3. Методы изучения Земли из космоса	19
1.3.1. Оптические методы	19
1.3.2. Радиотехнические методы	27
1.3.3. Спутниковые методы исследования атмосферы	33
1.4. Дистанционное зондирование Земли из космоса	34
1.4.1. Орбиты спутников	34
1.4.2. Прием спутниковой информации	41
1.4.3. Спутники для дистанционного зондирования	46
1.4.4. Искажения спутниковых сканерных изображений	55
Контрольные задания	59
Литература	60
2. Представление изображений в ЭВМ	62
2.1. Виды изображений	62
2.2. Датчики изображений	65
2.2.1. Передающая телевизионная трубка — видикон	65
2.2.2. Твердотельные датчики	66
2.3. Ввод изображений в ЭВМ	68
2.4. Устройства визуализации изображений	70
2.5. Представление изображений в ЭВМ	73
2.6. Алгоритмы сжатия изображений	74
2.7. Форматы графических файлов	78
2.8. Модели растровых изображений	81
2.8.1. Двухмерное случайное поле	82
2.8.2. Однородные и изотропные случайные поля	84
2.8.3. Модели случайных полей	89
2.8.4. Моделирование случайных полей в частотной области	96
Контрольные задания	98
Литература	99
3. Восстановление и улучшение изображений	100
3.1. Качество изображений	100
3.2. Предварительная обработка данных дистанционного зондирования	101
3.2.1. Геометрическая коррекция спутниковых изображений	101
3.2.2. Радиометрическая коррекция результатов дистанционного зондирования	104
3.2.3. Атмосферная коррекция	106
3.2.4. Восстановление пропущенных пикселей	109
3.2.5. Улучшение изображений путем изменения контраста	111
3.3. Линейная пространственно-инвариантная фильтрация	115
3.3.1. Линейные преобразования в частотной плоскости	117
3.3.2. Линейная фильтрация в частотной плоскости	120

3.3.3. Линейная локальная фильтрация	124
3.4. Нелинейная фильтрация изображений	131
3.4.1. Медианная фильтрация	131
3.4.2. Сигма-фильтр	134
3.4.3. Нелинейные градиентные фильтры	135
3.4.4. Кенстральная обработка	137
3.4.5. Ряд Вольтерра в нелинейной обработке изображений	139
Контрольные задания	143
Литература	143
4. Распознавание образов	145
4.1. Основы общей теории распознавания образов	145
4.1.1. Подходы к распознаванию изображений	147
4.2. Сегментация изображений	153
4.2.1. Пороговая сегментация и кластерный анализ	154
4.2.2. Сегментация путем наращивания областей	160
4.2.3. Сегментация путем выделения границ	163
4.3. Обнаружение объектов известной формы на изображении	164
4.4. Применение теоремы Байеса. Решающее правило	167
4.5. Параметрические методы классификации	171
4.5.1. Нормальный закон распределения в задачах классификации	171
4.5.2. Метод максимального правдоподобия	175
4.6. Простые методы, основанные на группировании	178
4.7. Непараметрические методы классификации	180
4.8. Анализ главных компонент	182
4.9. Применение нейрокompьютеров	187
Контрольные задания	190
Литература	191
5. Применение космических средств контроля окружающей среды	192
5.1. Географические информационные системы	192
5.1.1. ГИС мониторинга лесных пожаров	192
5.2. Проблема обнаружения лесных пожаров	207
5.2.1. Общая характеристика лесного пожара	207
5.2.2. Космический лесопожарный мониторинг	208
5.3. Контроль состояния растительности	215
5.4. Контроль состояния водоемов	221
5.5. Контроль снегового и ледового покрова	224
5.6. Исследование вертикальных профилей атмосферы	229
5.7. Изучение облачности	236
5.8. Проблема атмосферного озона	246
5.8.1. Ультрафиолетовое излучение Солнца	246
5.8.2. Экологическая роль озона	247
5.8.3. Образование и разрушение озона	248
5.8.4. Методы измерения общего содержания озона	251
5.8.5. Пространственно-временная динамика стратосферного озона	253
5.8.6. Циркумполярные вихри и озоновая дыра	255
5.9. Пакет IDRISI для цифровой обработки изображений	259
Контрольные задания	260
Литература	261

ПРЕДИСЛОВИЕ

В последнее десятилетие для России важное значение приобрели спутниковые методы исследования ее территории. Это связано как с дальнейшим совершенствованием космической техники, так и со свертыванием авиационных и наземных методов мониторинга. Особенно большое значение спутниковые методы имеют для севера страны и азиатской ее части.

Основные области применения спутникового дистанционного зондирования — получение информации о состоянии окружающей среды и землепользовании, изучение растительных сообществ, оценка урожая сельскохозяйственных культур, оценка последствий стихийных бедствий: наводнений, землетрясений, извержений вулканов, лесных пожаров. Средства дистанционного зондирования эффективны при изучении загрязнения почвы и водоемов, льдов на суше и на воде, в океанологии. Эти средства позволяют получать сведения о состоянии атмосферы, в том числе в глобальном масштабе. Данные зондирования поступают в виде изображений, как правило, в цифровой форме, обработка ведется на ЭВМ, поэтому проблематика дистанционного зондирования тесно связана с цифровой обработкой изображений.

К сожалению, в настоящее время ощущается нехватка научной и методической литературы по указанным вопросам. Этот вакуум могли бы восполнить переводы многочисленных зарубежных монографий и учебных пособий, однако, во-первых, таких переводов практически нет, а во-вторых, в зарубежных изданиях не отражена российская специфика. Эти причины и побудили авторов написать предлагаемую книгу.

Учебное пособие подготовлено на основе конспектов лекций, читаемых авторами, с использованием материалов практических занятий студентов Красноярского государственного технического университета (специальности 071903 «Геоинформационные системы», 320700 «Защита окружающей среды и рациональное природопользование») и Красноярского государственного университета (специальность 010400 «Физика»). При этом учтен опыт работы станции HRPT приема информации со спутников дистанционного зондирования Земли NOAA, переданных Российской академии наук Национальной администрацией США по авионавигации и исследованию космического пространства (NASA) в результате межправительственного соглашения между Россией и США. При этом исключительно важную роль сыграли академик А.С. Исаев (Международный институт леса, Москва) и доктор Р. Мерфи (NASA). Установленная в 1994 г. в Институте леса СО РАН им.

В.Н. Сукачева (Красноярск) станция НРРГ по своему техническому и программному оснащению является уникальной для России и стран СНГ. Ежедневно производится прием информации с нескольких спутников NOAA, накоплен обширный архив спутниковых данных о Сибири. Российским коллективом сотрудников станции при участии аспирантов и студентов КГУ и КГТУ разработаны оригинальные методики и программные продукты.

Глава 1 написана совместно В.Б. Кашкиным и А.И. Сухининым; главы 2, 4 — В.Б. Кашкиным; §5.1–5.7 — А.И. Сухининым, §5.8 — А.И. Сухининым и А.В. Черепановым, остальное — В.Б. Кашкиным.

1. КОСМИЧЕСКИЕ СРЕДСТВА КОНТРОЛЯ ОКРУЖАЮЩЕЙ СРЕДЫ

1.1. ОБЩАЯ ХАРАКТЕРИСТИКА ДИСТАНЦИОННЫХ МЕТОДОВ

Для наблюдения Земли из космоса используют дистанционные методы: исследователь имеет возможность на расстоянии получать информацию об изучаемом объекте. Дистанционные методы, как правило, являются косвенными, т.е. с их помощью измеряют не интересующие нас параметры объектов, а некоторые связанные с ними величины. Например, нам необходимо оценить состояние сельскохозяйственных посевов. Но аппаратура спутника регистрирует лишь интенсивность светового потока от этих объектов в нескольких участках оптического диапазона. Чтобы «расшифровать» такие данные, требуются предварительные исследования, включающие в себя различные эксперименты по изучению состояния растений контактными методами; по изучению отражательной способности листьев в различных участках спектра и при различном взаимном расположении источника света (Солнца), листьев и измерительного прибора. Далее необходимо определить, как выглядят те же объекты с самолета, и лишь после этого судить о состоянии посевов по спутниковым данным.

Методы изучения Земли из космоса не случайно относят к высоким технологиям. Это связано не только с использованием ракетной техники, сложных оптико-электронных приборов, компьютеров, но и с новым подходом к получению и интерпретации результатов измерений. И хотя трудоемкие подспутниковые исследования проводятся на небольшой площади, они дают возможность обобщать данные на огромные пространства и даже на весь земной шар. Широта охвата является характерной чертой спутниковых методов исследования Земли. К тому же эти методы, как правило, позволяют получать результат за сравнительно короткий интервал времени. В настоящее время для Сибири с ее бескрайними просторами спутниковые методы единственно приемлемы.

Примеры изображений Земли из космоса представлены на рис. 1.1 и 1.2.

К числу особенностей дистанционных методов относится влияние среды (атмосферы), через которую проходит сигнал со спутника. Самый простой пример такого влияния — наличие облачности, закрывающей интересующие объекты и делающей невозможным наблюдения в оптическом диапазоне. Однако и при отсутствии облачности атмосфера ос-

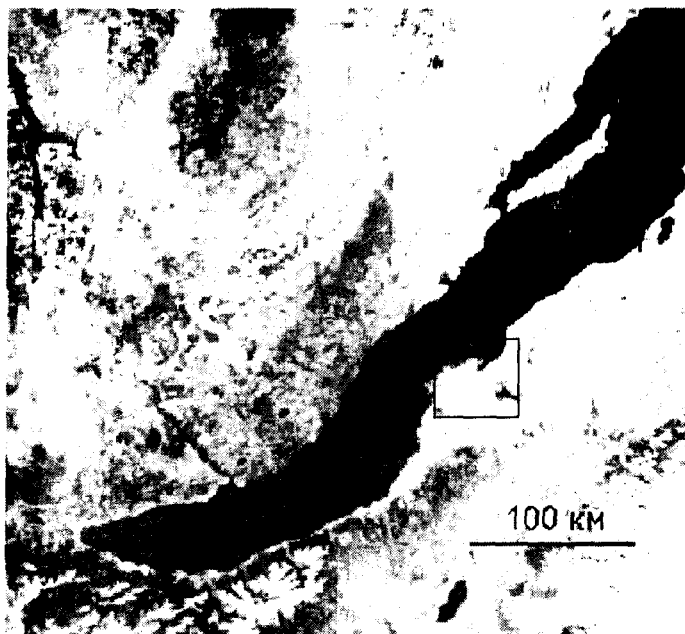


Рис. 1.1. Изображение озера Байкал со спутника NOAA (разрешение 1,1 км)

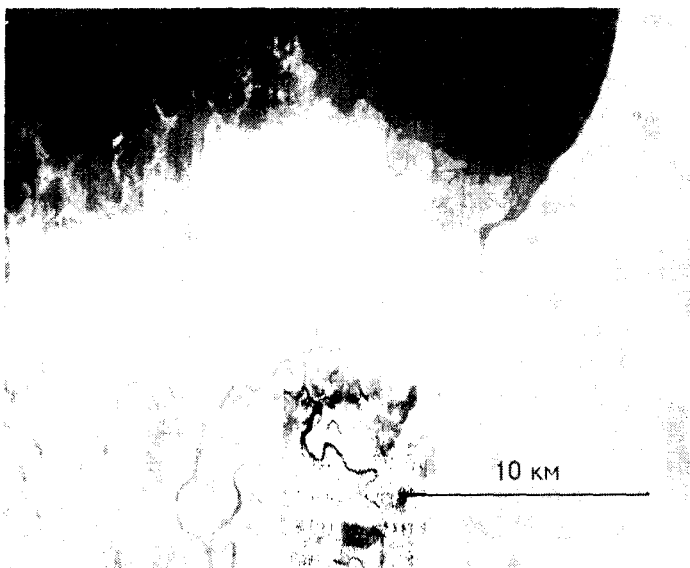


Рис. 1.2. Место впадения р. Селенги в озеро Байкал (разрешение 45 м)

лабляет излучение от объекта, особенно в полосах поглощения составляющих ее газов. Поэтому приходится работать в так называемых окнах прозрачности, учитывая, что и в них имеют место поглощение и рассеяние излучения газами и аэрозолями. В радиодиапазоне возможно наблюдение Земли и сквозь облачность.

Информация о Земле поступает со спутников, как правило, в цифровом виде, что также характерно для дистанционных методов. Наземная цифровая обработка изображений проводится на ЭВМ; в настоящее время она относится к числу наиболее динамично развивающихся информационных технологий, применяемых в робототехнике, полиграфии, медицине, физическом материаловедении и т.д.

Современные спутниковые методы позволяют не только получать изображение Земли. Используя чувствительные приборы, удается измерять концентрацию атмосферных газов, в том числе вызывающих парниковый эффект. Спутник «Метеор-3» с установленным на нем прибором TOMS позволял за сутки оценить состояние всего озонового слоя Земли. Спутник NOAA кроме получения изображений поверхности дает возможность исследовать озоновый слой и даже изучать вертикальные профили параметров атмосферы (давление, температуру, влажность на разных высотах в сотнях точек в полосе обзора).

Дистанционные методы делят на активные и пассивные. При использовании активных методов спутник посылает на Землю сигнал собственного источника энергии (лазера, радиолокационного передатчика), регистрирует его отражение. Радиолокация позволяет «видеть» Землю сквозь облака. Чаще используются пассивные методы, когда регистрируется отраженная поверхностью энергия Солнца либо тепловое излучение Земли.

1.2. ФИЗИЧЕСКИЕ ОСНОВЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

1.2.1. Излучение энергии нагретыми телами

При дистанционном зондировании Земли из космоса используются оптический диапазон электромагнитных волн и микроволновый участок радиодиапазона. Оптический диапазон (рис. 1.3) включает в себя ультрафиолетовый (УФ) участок спектра; видимый участок — синюю (В), зеленую (Г) и красную (Р) полосы; инфракрасный участок (ИК) — ближний (БИК), средний и тепловой.

При пассивных методах зондирования в оптическом диапазоне источниками электромагнитной энергии являются разогретые до достаточно высокой температуры твердые, жидкие, газообразные тела.

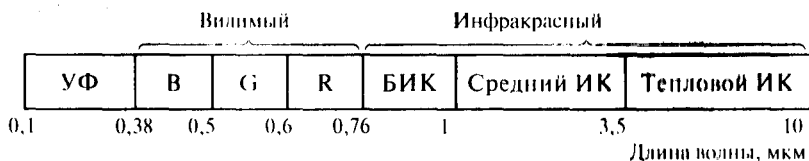


Рис. 1.3. Оптический диапазон электромагнитных волн

При термодинамическом равновесии с окружающей средой все тела с одинаковой температурой T излучают одинаково (первый закон Кирхгофа). При этом энергия, поглощаемая участком поверхности в секунду, равна энергии, излучаемой в тот же промежуток времени той же поверхностью (второй закон Кирхгофа).

Интенсивность излучения I в заданном направлении, характеризуемом углом α между нормалью и излучающей поверхностью абсолютно черного тела, определяется законом Ламберта: $I = I_0 \cos \alpha$, где I_0 — максимальная интенсивность излучения при $\alpha = 0$; при $\alpha = 90^\circ$, т.е. по касательной к поверхности, интенсивность излучения равна нулю.

По формуле Планка поток энергии, излучаемой в состоянии термодинамического равновесия единицей поверхности абсолютно черного тела с температурой T в интервале длин волн λ , $\lambda + d\lambda$,

$$B(\lambda, T) = \frac{c_1}{\lambda^5} \frac{1}{\left[\exp\left(\frac{c_2}{\lambda T}\right) - 1 \right]}. \quad (1.1)$$

Здесь $c_1 = 3,7415 \cdot 10^8$ Вт·мкм⁴/м²; $c_2 = 14\,388$ мкм · К. Максимум излучения приходится на $\lambda = 2898/T$ мкм.

Полная энергия во всем интервале длин волн описывается формулой Стефана–Больцмана:

$$\int_0^\infty B(\lambda, T) d\lambda = a \cdot T^4,$$

где $a = 5,67 \cdot 10^{-8}$ Вт · м⁻² · К⁻⁴.

При наблюдении Земли из космоса при длине волны короче 2–3 мкм регистрируется энергия Солнца, отраженная и рассеянная поверхностью суши, воды и облаков. Температура поверхности (фотосферы) Солнца равна 5785 К, максимум излучения приходится на 0,5 мкм (рис. 1.4). Озон, в небольшом количестве содержащийся в ат-

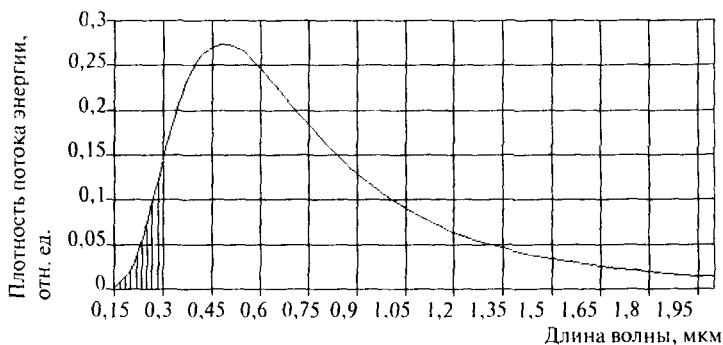


Рис. 1.4. Распределение энергии в спектре Солнца согласно закону Планка

мосфере, сильно поглощает опасно воздействующее на животный и растительный мир ультрафиолетовое излучение с длиной волны короче 0,3 мкм, так что при наблюдении Солнца с поверхности Земли отсутствует коротковолновый скат кривой $B(\lambda, T)$ (на рис. 1.4 заштрихован). В остальном реальный спектр Солнца мало отличается от рис. 1.4.

Глаз человека видит предметы в интервале длин волн 0,38–0,76 мкм, максимум чувствительности приходится на $\lambda = 0,5$ мкм.

На волнах длиной более 4 мкм собственное тепловое излучение Земли превосходит излучение Солнца. Регистрируя интенсивность теплового излучения Земли из космоса, можно достаточно точно оценивать температуру суши и водной поверхности, которая является важнейшей экологической характеристикой. Измерив температуру верхней границы облачности (ВГО), можно определить ее высоту, если учесть, что в тропосфере с высотой температура уменьшается в среднем на $6,5^\circ/\text{км}$.

При регистрации теплового излучения со спутников используется интервал длин волн 10–14 мкм, в котором поглощение в атмосфере невелико. При температуре земной поверхности (облаков), равной -50°C , максимум излучения согласно (1.1) приходится на 12 мкм, при $+50^\circ\text{C}$ — на 9 мкм.

Если с помощью датчика, установленного на спутнике, измерен поток энергии мощности $B = B(\lambda, T)$ от некоторого объекта, то из (1.1) получаем $T = (\lambda/c_2) \ln(c_1/\lambda^5 B + 1)$. Определенная по интенсивности B теплового излучения (радиации) температура T называется *радиационной*, в отличие от *термодинамической* температуры, характеризующей интенсивность теплового движения молекул вещества и измеряемой контактным термометром. Для абсолютно черного тела обе

температуры совпадают, для реальных тел — нет, так как для них величина B составляет некоторую долю ϵ от плотности потока мощности, излучаемой абсолютно черным телом при той же температуре; ϵ носит название *коэффициента теплового излучения*. В тепловом ИК-диапазоне коэффициент теплового излучения ближе всего к 1, для воды 0,98–0,99. Для облаков также $\epsilon \sim 1$. Сложнее дело обстоит с поверхностью суши. Здесь проявляется характер поверхности (гладкая или нет), ее увлажненность и т.д. Для свежего снега $\epsilon = 0,986$, для густой травы 0,970, глинистой почвы 0,980, хвойного леса 0,97. Для достаточно точного (с погрешностью не более 0,2–0,5 К) определения температуры поверхности из космоса необходимо учитывать и коэффициент теплового излучения, и поглощение в атмосфере.

Пример определения температуры воды со спутника в предположении, что для водной поверхности $\epsilon = 1$, показан на рис. 1.5.

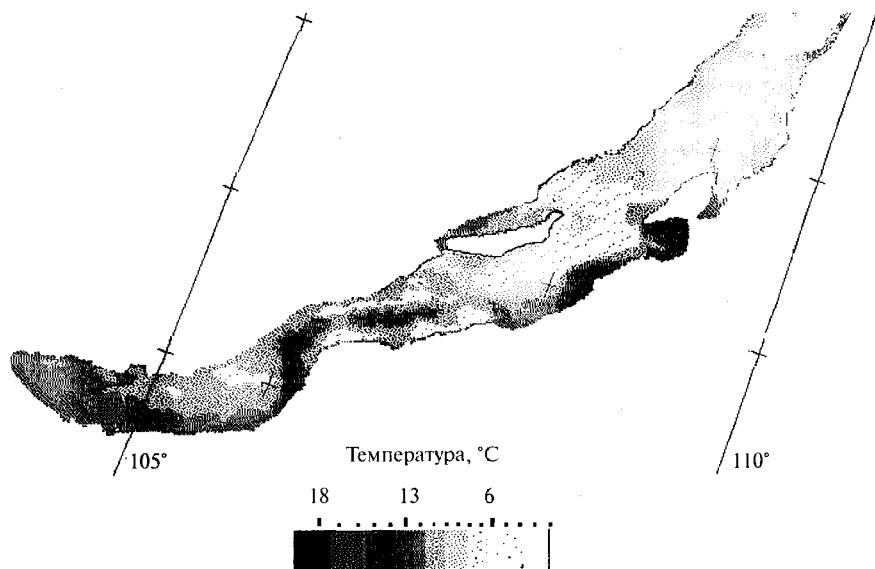


Рис. 1.5. Температурная карта озера Байкал по данным спутника NOAA-14 за 12 июля 1998 г.

1.2.2. Отражение и рассеяние солнечной энергии поверхностью

Установленная на спутнике аппаратура в видимом и ближнем ИК-диапазоне регистрирует солнечную энергию, отраженную и рассеянную поверхностью Земли. Отражательная способность A — это отношение

количества (плотности потока) отраженной и рассеянной вверх радиации I_0 к плотности потока падающей прямой радиации Солнца I_c : $A = I_0/I_c$. Обычно A выражают в % и говорят об общей отражательной способности для широкой области видимого и ближнего инфракрасного спектра. Можно определять A и для отдельных участков спектра. И вообще, значение A для различных типов поверхности и различных участков спектра — это ключ к распознаванию деталей на спутниковых изображениях Земли.

Для идеально матовой поверхности зависимость интенсивности радиации, рассеянной под углом α к нормали к поверхности, в точности подчиняется закону Ламберта. Такую поверхность имеют, например, облака. Для других типов поверхности закон Ламберта выполняется приближенно.

Значение A зависит от свойств (в том числе от влажности) отражающей поверхности (трава, лес, почва и т.п.), от спектра падающей радиации — прямой и рассеянной, от угла ее падения ζ , рельефа и т.п. Так, для зеленой растительности, образующей хорошо поглощающий слой, A мало в видимой части спектра (около 5%) и велико в ближней ИК-области. У оголенных почв отражательная способность очень различна: у подзола — всего 7%, у сухого песка в красной части спектра — до 37% (и заметно меньше у влажного песка). И у растений, и у почвы A обычно растет с увеличением ζ (при низком Солнце). В зимнем хвойном лесу снег, лежащий между деревьями, увеличивает A при малом ζ и не влияет на A при косых солнечных лучах.

Для свежего снега A достигает 98% в видимых лучах и гораздо меньше в ИК-области. У старого зернистого и влажного снега и морского льда $A \approx 43\text{--}45\%$. При $\lambda = 1,55\text{--}1,65$ мкм снег почти черный ($A \sim 0$), что при спутниковых наблюдениях позволяет отличать заснеженные участки от облаков, для которых на этих длинах волн A велико.

Для поверхности водоемов в видимой области $A \approx 4\text{--}6\%$ при малых ζ и еще меньше в ИК-спектре с $\lambda = 0,8\text{--}1,5$ мкм, но возрастает до 30% и выше при большом угле падения лучей или при зеркальном их отражении ($\zeta > 60^\circ$), когда, например, со спутника видна светлая дорожка солнечных бликов на воде. Волнение увеличивает A до 13% при малых ζ и уменьшает — при больших. Часть отражательной способности (1–4%) обусловлена светом, рассеянным в верхнем тонком слое воды.

Среднее значение A поверхности Земли меньше 20% в широком поясе между 55° с. ш. и 60° ю. ш., особенно в океанических областях. Весь год оно выше 80% в Антарктике. В северной околополярной области и в области снежного покрова, которая зимой местами распространя-

ется к югу до 40° с. ш., оно меняется с сезоном, но также велико (65–75%).

Отражательная способность облаков, наблюдаемых сверху, тем больше, чем мощнее их слой. Так, слой слоистых облаков толщиной 300 м имеет $A = 59$ –63%.

Особый интерес представляет отражательная способность растительного покрова, которая определяется пигментами листа, особенно хлорофиллом (рис. 1.6). В синей и красной областях видимого диапа-



Рис. 1.6. Отражательная способность листьев сои (1) и кукурузы (2) с одинаковым содержанием влаги (66–80%)

зона отражательная способность листа низкая из-за наличия полос поглощения с центрами при $\lambda_1 = 0,45$ и $\lambda_2 = 0,65$ мкм. Между полосами, т.е. в зеленой области, максимум отражательной способности лежит приблизительно при 0,54 мкм. При этом нормальная, здоровая листва воспринимается глазом как зеленая. Когда растение находится в состоянии стресса и образование хлорофилла уменьшается, поглощение энергии в полосах λ_1 и λ_2 уменьшается, отражательная способность повышается, что особенно заметно в красной области. Поэтому растение кажется желтоватым.

В ближнем ИК-диапазоне при $\lambda > 0,75$ мкм поглощательная способность зеленого листа мала (~5%), отражательная способность достаточно велика (до 45–50%), значительное количество энергии проходит сквозь лист (45–50%). Отражательная способность нескольких слоев листьев выше (до 85%), чем одного листа. В среднем ИК-диапазоне вода имеет несколько полос поглощения (при $\lambda = 1,4$; 1,9 и 2,7 мкм), в которых спектральный отклик зеленых листьев ослаблен. У хвойного леса для всего видимого диапазона $A = 6$ –16%, у лиственного леса летом $A = 7$ –18%, у травы $A = 7$ –25%, у поля близких к зрелости злаков $A = 13$ –23%.

Установленная на спутнике NOAA аппаратура регистрирует восходящее от земной поверхности излучение в красном (0,58–0,68 мкм, полоса поглощения λ_2), ближнем инфракрасном (0,725–1,0 мкм) и некоторых других областях спектра. Сравнивая между собой яркости f_1 и f_2 изображения растительности в красном и ближнем ИК-участках, можно судить о состоянии растений в данном районе. Для этого используют так называемый дифференциальный вегетационный индекс $DVI = f_2 - f_1$. Однако DVI очень чувствителен к взаимному расположению источника освещения (Солнца), спутника и растений. Нормализованный дифференциальный вегетационный индекс

$$NDVI = \frac{f_2 - f_1}{f_1 + f_2} \quad (1.2)$$

в значительной степени лишен этого недостатка. Это связано с тем, что f_1 и f_2 изменяются приблизительно пропорционально при изменении взаимного расположения Солнца, спутника и растений. При обработке данных со спутника NOAA нередко используют не $NDVI$, а половину этой величины.

Для растений в нормальном состоянии $NDVI$ близок к 0,6–0,65; низкие значения $NDVI$ (порядка 0,3–0,4) свидетельствуют о недостатке влаги или о заболевании растения. Угнетение вегетации может быть вызвано неблагоприятной экологической обстановкой, поэтому $NDVI$ является также индикатором загрязнения окружающей среды.

Более точно о состоянии растительности из космоса можно судить, если использовать не два, а большее число участков видимого и ближнего ИК-диапазона. В отличие от спутника NOAA спутники «Ресурс-О1-3», LANDSAT и SPOT позволяют дополнительно измерять яркость объектов в зеленом (0,5–0,6 мкм) участке спектра, т.е. точнее различать сельскохозяйственные культуры, определять занимаемую ими площадь и оценивать урожайность. Спутники с многоспектральной аппаратурой эффективны при мониторинге загрязнения окружающей среды, при поиске полезных ископаемых, при изучении океанских ресурсов и др.

1.2.3. Влияние атмосферы

При дистанционном зондировании поверхности Земли атмосфера является возмущающей средой, которая искажает спутниковые данные, а в некоторых участках электромагнитного диапазона, например в дальнем инфракрасном с длиной волны около 100 мкм, вообще препятст-

вует дистанционному зондированию. С другой стороны, спектральные линии поглощения газов однозначно характеризуют эти газы, а интенсивность и ширина линий отражают их физические параметры (температуру, плотность, общее количество молекул). Поэтому спектральные линии очень важны при дистанционном зондировании самой атмосферы.

До высоты 100 км атмосферные газы равномерно перемешаны. К основным составляющим атмосферы относят кислород O_2 (около 21% воздуха по объему), азот N_2 (около 78%) и аргон Ar (несколько менее 1%). Влияние их на наблюдение Земли из космоса незначительно.

Важным компонентом атмосферы является водяной пар, содержание которого в атмосфере непостоянно и относительно невелико. Он имеет очень большое число линий поглощения в инфракрасном и микроволновом диапазонах спектра, начиная с $\lambda = 0,72$ мкм и далее у 0,81; 0,94; 1,1; 1,38; 1,87; 2,7–3,2; 6,3 мкм. Широкая линия поглощения перекрывает диапазон длин волн примерно от 10 мкм до 1 см.

Углекислый газ CO_2 имеет две узкие линии поглощения при $\lambda = 2,7$ мкм и 4,26 мкм и вызывает сильное поглощение в дальней инфракрасной зоне спектра, начиная с $\lambda = 13$ мкм.

Известно, что в атмосфере в небольших количествах содержится 20–30 так называемых малых газов, имеющих как естественное, так и антропогенное происхождение. Некоторые из них, как и углекислый газ, могут влиять на климат Земли и на здоровье человека. К малым газам естественного происхождения относят закись азота N_2O и метан CH_4 (результат деятельности бактерий), сернистый ангидрид SO_2 , сернистый карбонил COS , сероводород H_2S и др. (вулканические выбросы), а также озон O_3 , возникающий в верхних слоях атмосферы под воздействием ультрафиолетового излучения Солнца. К малым газам антропогенного происхождения причисляют отходы топок, производства, транспорта, сельского хозяйства: SO_2 , CO , хлороводород HCl , фреоны CCl_xF_y и др.

Естественный аэрозоль — твердые и жидкие частицы, взвешенные в воздухе, — включает космическую, вулканическую и почвенную минеральную пыль, пыльцу растений, частицы морской соли, капли облаков и туманов, частицы дыма лесных и торфяных пожаров. Антропогенный аэрозоль — это частицы сажи, пепла, цемента и других отходов производства и транспорта. Существенным источником загрязнения являются топки печей, ГРЭС, тепловые двигатели.

Для многоатомных малых газов, а также CO_2 характерны электронные переходы с энергией диссоциации порядка 3–5 эВ, а также вращательно-колебательные переходы с энергией возбуждения порядка 0,1 эВ

и менее. Первые из них связаны с солнечным излучением УФ- и видимого диапазонов, вторые — с квантами ИК- и радиодиапазона.

Молекулы озона O_3 , находящиеся в основном в стратосфере, сильно влияют на общее поглощение только в одном небольшом участке инфракрасной зоны ($\lambda = 9,59$ мкм). Однако озон активно поглощает энергию в ультрафиолетовой зоне и в зоне миллиметровых радиоволн. Например, слой озона толщиной всего 3 мм (при нормальном давлении и температуре) в 10^{40} раз ослабляет УФ-излучение Солнца длиной волны $\lambda = 0,255$ мкм. И озон, и другие газы при поглощении преобразуют энергию Солнца в тепловую.

Малые газы, водяной пар и CO_2 вызывают ослабление излучения из-за дискретного поглощения в таких важных диапазонах длин волн, как ближний инфракрасный (~ 1 мкм) и тепловой инфракрасный (> 10 мкм). В результате в дальнем ИК остается прозрачным только один широкий диапазон длин волн 8–12 мкм, а в ближнем и среднем ИК в зоне длин волн менее 4 мкм имеются четыре узких диапазона, которые и используются для дистанционного зондирования. Прозрачен и видимый диапазон, однако в наиболее коротковолновой его части (фиолетовый и голубой участки) велики «помехи» от молекулярного рассеяния солнечного света.

Атмосфера не прозрачна в дальней ИК-зоне спектра, где длина волны более 14 мкм. И только при волне длиной около 1 мм атмосфера вновь становится более прозрачной, поскольку лишь некоторые, более слабые переходы вызывают поглощение (рис. 1.7). Ослабление атмосферой восходящего излучения от поверхности Земли в полосах прозрачности описывается объемным коэффициентом поглощения α , а рассеяние света молекулами газов, капельками воды, пылинками — объемным коэффициентом рассеяния σ . Оптическая толщина τ — это произведение объемного коэффициента ослабления света атмосферой на геометрическую длину пути светового луча. Если коэффициенты α и σ зависят от координаты (а плотность атмосферы меняется с высотой), то на пути между точками с координатами r_1 и r_2



Рис. 1.7. Окна прозрачности атмосферы в оптическом диапазоне

$$\tau(r_1, r_2) = \int_{r_2}^{r_1} |\alpha(r) + \sigma(r)| dr.$$

При малых эффектах многократного рассеяния, т.е. в пределах справедливости закона Бугера, $I_1 = I_0 \exp(-\tau)$, где I_0 и I_1 — интенсивности в начальной (на поверхности Земли) и в конечной (на орбите спутника) точках при наблюдении в нади́р. Эффектами многократного рассеяния можно пренебречь для зеленого (G) и более длинноволновых участков спектра, но не для синего (B).

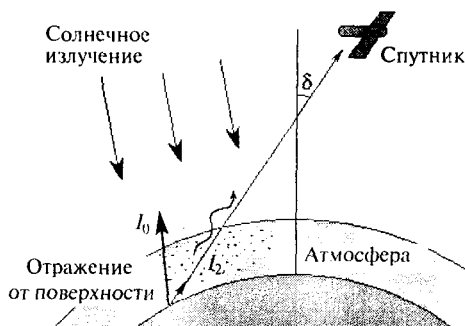


Рис. 1.8. Диаграмма, поясняющая наблюдение под углом к нади́ру

Приведенная формула справедлива для наблюдения в нади́ре. Если поверхность наблюдается под углом δ к нади́ру, то путь, пройденный излучением в атмосфере, увеличится (рис. 1.8). В этом случае спутник регистрирует излучение $I_2 = I_0 \exp(-\tau \sec \delta)$. Формула не учитывает кривизну Земли; существуют методики, позволяющие провести такой учет. Не учитывается также

влияние возможных горизонтальных локальных неоднородностей (наличие тумана, дымки, облаков пыли и т.д.), поэтому расчет приближенный.

В атмосфере рассеяние света молекулами различных веществ подчиняется некоторым закономерностям. Интенсивность молекулярного рассеяния пропорциональна λ^{-4} , что наиболее заметно в коротковолновой части спектра, ответственной за голубой цвет неба. Рассеяние на аэрозолях (размер частиц от 0,1 до 10λ) пропорционально $\lambda^{-\alpha}$, $0 < \alpha < 4$. Размеры молекул дыма и облаков много больше длины волны видимого и ИК-диапазонов, поэтому интенсивность рассеянного солнечного излучения не зависит от длины волны и цвет дымов и облаков близок к белому.

Спутник может принимать информацию с больших площадей, при этом различные участки поверхности наблюдаются под разными углами δ и луч от них проходит разные расстояния в атмосфере. В результате яркость одинаковых по оптическим свойствам участков поверхности будет различной. Для расчетов по формуле Бугера необходимо знать оптическую толщину атмосферы τ , для чего на Земле следует из-

мерять ослабление солнечной радиации. Значение τ можно определять со спутника, наблюдая яркость объектов с заранее известными оптическими характеристиками. Можно также грубо оценить τ в видимом участке спектра, используя зависимость между τ и метеорологическими характеристиками тропосферы: температурой воздуха, скоростью ветра, относительной влажностью и упругостью водяного пара.

Более точно учесть влияние атмосферы можно на основе теории переноса излучения в атмосфере, которая позволяет оценить яркость объектов по данным измерения восходящего излучения путем решения обратной задачи.

1.3. МЕТОДЫ ИЗУЧЕНИЯ ЗЕМЛИ ИЗ КОСМОСА

1.3.1. Оптические методы

Первые изображения Земли из космоса были получены с помощью фотокамеры. Эта методика применяется и в настоящее время. Спутник с фоторегистацией «Ресурс-Ф1М» (Россия) позволяет фотографировать Землю в интервале длин волн 0,4–0,9 мкм. Отснятые материалы спускаются на Землю и проявляются. Анализ снимков, как правило, проводится визуально с помощью проекционной аппаратуры, которая позволяет также получать цветные фотоотпечатки. Метод обеспечивает высокую геометрическую точность изображения; можно увеличить снимки без заметного ухудшения качества. Однако он малооперативен, поскольку изображение представлено в виде фотографий, а не в цифровой форме, и эффективен в видимом и ближнем ИК-диапазонах.

Этих недостатков лишены сканерные методы. Сканер с цилиндрической разверткой в принципе представляет собой маятник, закрепленный в одной точке и колеблющийся поперек направления движения аппарата (рис. 1.9). На конце маятника в его фокальной плоскости установлен объектив с точечным фотоприем-



Рис. 1.9. Схема сканирования поверхности Земли

ным устройством (фотоэлектронный умножитель, фотодиод, фоторезистор). При движении аппарата над Землей с выхода фотоприемного устройства снимается сигнал, пропорциональный освещенности в видимом или ближнем ИК-диапазоне того участка земной поверхности, на который в данный момент направлена ось объектива. Если фотоприемное устройство — фоторезистор, то можно регистрировать излучение в тепловом ИК-диапазоне и определять температуру поверхности и облаков. На практике сканер неподвижен, а качается (вращается) зеркало, отражение от которого через объектив попадает на фотоприемное устройство. Сканерная информация в цифровой форме передается со спутника в реальном времени или в записи на бортовой магнитофон, на Земле она обрабатывается на ЭВМ.

Линейный сканер содержит расположенные в линию неподвижные фоточувствительные элементы 190—1000 и более на приборах с зарядовой связью (ПЗС) — линейку ПЗС или несколько таких линеек длиной порядка сантиметра. На линейки через объектив фокусируется изображение земной поверхности, все элементы находятся в фокальной плоскости. Линейка, ориентированная поперек направления движения спутника, перемещается вместе с ним, последовательно «считывая» сигнал, пропорциональный освещенности различных участков поверхности и облаков. Линейные сканеры на ПЗС работают в видимом и ближнем ИК-диапазонах.

Сканер МСУ-СК, устанавливаемый на российских спутниках «Ресурс-О» и др., единственный, в котором реализован перспективный принцип конической развертки, заключающийся в перемещении визирного луча по поверхности конуса с осью, направленной в нади́р. Сканирующий луч описывает по сферической поверхности Земли дугу (обычно в переднем секторе сканирования). За счет перемещения спутника изображение представляет собой совокупность дуг. Достоинством такого вида развертки является постоянство угла между поверхностью Земли и направлением на спутник, что особенно важно при изучении растительности. Постоянно также расстояние L от спутника до каждой точки дуги, так что разрешение сканера МСУ-СК, в отличие от сканеров с цилиндрической и линейной разверткой, постоянно по всему изображению. При этом для достаточно больших участков изображения постоянно и атмосферное ослабление восходящего излучения и нет необходимости в атмосферной коррекции. Отсутствуют также искажения изображения за счет кривизны Земли, характерные для других сканеров.

Важными характеристиками сканера являются полоса обзора G , радиометрическая разрешающая способность ΔI , мгновенное поле зрения Δj и разрешение на местности ΔL . Радиометрическая разрешающая

гноспособность (яркостная, температурная) определяется шириной динамического диапазона датчика и числом уровней дискретизации (числом бит), соответствующих переходу от яркости абсолютно «черного» к абсолютно «белому».

Мгновенное поле зрения зависит от характеристик объектива и размеров фотоприемника. Это интервал углов $\Delta\varphi$, в пределах которых в каждый момент времени на фотоприемник попадает изображение земной поверхности или облаков. Если спутник находится на расстоянии L от некоторой области на поверхности Земли, то фотоприемник одновременно регистрирует излучение от участка поверхности с линейным размером $\Delta L = \Delta\varphi \cdot L$ в пределах мгновенного угла зрения (здесь $\Delta\varphi$ выражено в радианах). Величина ΔL — это разрешение сканера на местности. Если два объекта на поверхности Земли находятся на расстоянии менее ΔL , то их изображения, вообще говоря, сливаются. Сканерное изображение обычно состоит из отдельных элементов квадратной или прямоугольной формы с линейным размером ΔL — пикселей, которые хорошо видны при соответствующем увеличении изображения. Не только спутниковые, но и все другие цифровые изображения состоят из отдельных пикселей.

Важная характеристика при работе с картой — ее масштаб. Если 1 см карты соответствует 1 км на местности, то масштаб равен 1:100 000, так как в 1 км содержится 100 000 см. Разрешение на местности, даваемое спутником, — это другая характеристика, но бывает необходимо сравнить спутниковое изображение и карту некоторого масштаба. В пределах пиксела спутниковое изображение однородно. На 1 см карты, как и на изображении, содержащем приблизительно 10 пикселей, можно рассмотреть некоторые детали. Поэтому условно можно считать, что разрешение в 1 км соответствует масштабу карты в 1:1 000 000, разрешение в 100 м — масштабу 1:100 000, разрешение в 30 м — масштабу 1:30 000.

Пространственное разрешение ΔL зависит от параметров оптической системы и от расстояния L от спутника до объекта: ΔL — наилучшее (минимальное) в подспутниковой точке (надире), когда расстояние L минимально и равно высоте спутника над Землей. При отклонении от надира у сканеров с цилиндрической и линейной разброской L увеличивается и разрешение становится хуже. Разрешение сканера AVHRR спутника NOAA в надире $\Delta L = 1,1$ км, при максимальном отклонении сканера от направления в надир разрешение становится равным приблизительно 4 км. У сканера МСУ-Э спутника «Ресурс-О» $\Delta L = 45$ м в надире.

В объективах сканеров используется, как правило, зеркальная оптика, а не линзовая, поскольку показатель преломления линз и поглощение света в них различны для различных длин волн, а сканер должен

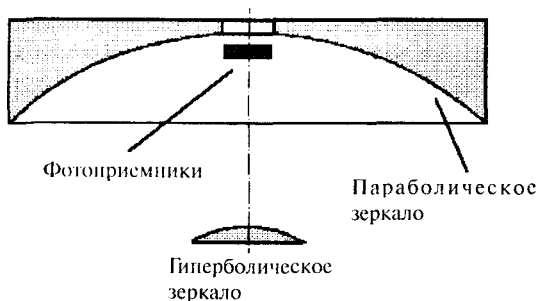


Рис. 1.10. Оптическая схема Кассегрена

работать в широком диапазоне электромагнитных волн — от видимого участка до инфракрасного. Зеркальные объективы имеют вогнутое зеркало параболической формы, на внутреннюю поверхность которого нанесена тонкая отражающая металлическая пленка

Такие объективы чаще всего строятся по схеме Кассегрена, когда перед основным параболическим зеркалом помещено второе небольшое выпуклое зеркало гиперболической формы (рис. 1.10). Свет, отраженный основным зеркалом, падает на вспомогательное и от него на площадку, где в фокальной плоскости объектива размещены фотоприемники. По своим свойствам двухзеркальный объектив равноценен однозеркальному, но с большим фокусным расстоянием.

Мгновенное поле зрения $\Delta\varphi$ определяется размером апертуры объектива и размерами фотоприемника. *Апертура* (действующее отверстие оптической системы) для параболического зеркала — это круг диаметром AB , замыкающий зеркало. В первом приближении направленные свойства объектива можно определять из следующих соображений. Пусть электромагнитное излучение от точечного источника M с координатами x_1, y_1, z_1 на земной поверхности (рис. 1.11) попадает на круглую апертурную площадку, расположенную на участке плоскости $CDEF$. Начало координат O , расположенное в центре круга, удалено от источника M на расстояние L в сотни и тысячи километров, тогда как размер апертурной площадки в оптическом диапазоне составляет десятки сантиметров (в радиодиапазоне — порядка 10–15 м). Поэтому можно утверждать, что амплитуда волны во всех точках объектива одинакова, хотя фаза волны разная. В точке B апертурной площадки с координатами x, y, z (здесь $z=0$) вектор электрического поля может быть представлен в виде

$$E = E_0 \exp i(\omega t - kL_B), \quad (1.3)$$

где угловая частота ω связана со скоростью света c и длиной волны λ соотношением $\omega = 2\pi c/\lambda$, $k = 2\pi/\lambda$ — волновое число, $L_B = [(x - x_1)^2 + (y - y_1)^2 + z_1^2]^{1/2}$ — расстояние между точками M и B .

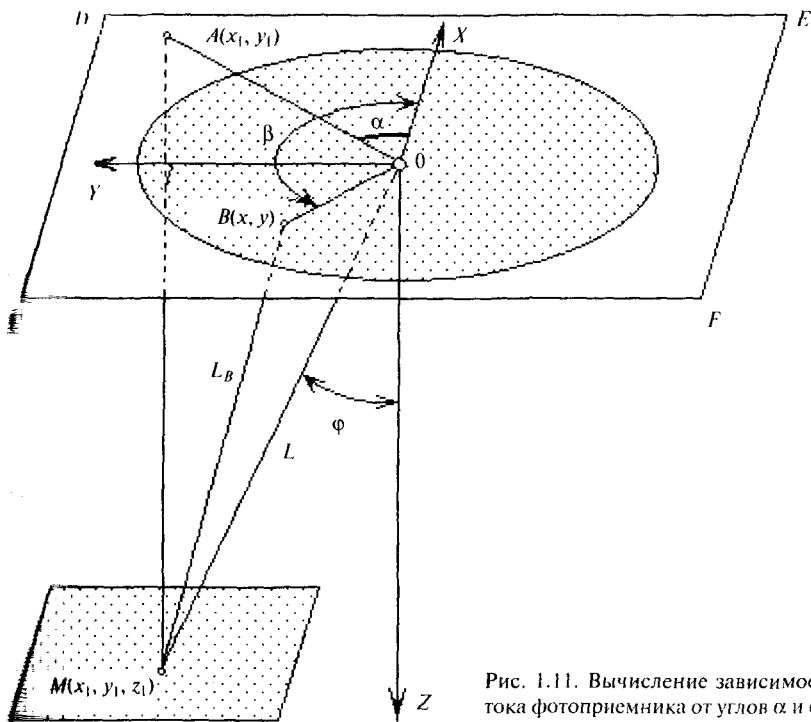


Рис. 1.11. Вычисление зависимости тока фотоприемника от углов α и ϕ

На рис. 1.11 из M на участок плоскости $CDEF$ в точку A опущен перпендикуляр длиной $z_1 = L \cos \phi$. Обозначим $x_1 = L \sin \phi \cos \alpha$, $y_1 = L \sin \phi \sin \alpha$, $L_B = L_B(L, x, y, \phi, \alpha) = [L^2 + x^2 + y^2 - 2xx_1 - 2yy_1]^{1/2}$. Но $x/L \ll 1$, $y/L \ll 1$, поэтому членами с x^2 и y^2 можно пренебречь. Электромагнитное излучение от всех точек апертурной площадки суммируется в фотоприемнике.

Чтобы найти зависимость тока фотоприемника от углов α и ϕ , необходимо проинтегрировать выражение для L_B по всем значениям x , y в пределах апертурной площадки. Поскольку эта площадка имеет круглую форму, имеет смысл перейти к полярным координатам: $x = \rho \cos \beta$, $y = \rho \sin \beta$, где $\rho = (x^2 + y^2)^{1/2}$. Таким образом,

$$L_B = L [1 - (2\rho/L) \cos \beta \sin \phi \cos \alpha - (2\rho/L) \sin \beta \sin \phi \sin \alpha]^{1/2} = L [1 - (2\rho/L) \sin \phi \cos(\alpha - \beta)]^{1/2}. \quad (1.4)$$

Учитывая малость отношения ρ/L , разлагая (1.4) в ряд и принимая во внимание первые два члена, получаем $L_B = L - \rho \sin \phi \cos(\alpha - \beta)$.

Ток фотоприемника

$$I = \int_0^{2\pi} \int_0^r \exp[-jk\rho \sin\varphi \cos(\alpha - \beta)] \rho d\alpha d\rho, \quad (1.5)$$

где r — радиус апертурной площадки, $\rho d\alpha d\rho$ — якобиан преобразования декартовых координат к полярным. Интеграл по α в (1.5) равен $2\pi J_0(k\rho \sin\varphi)$, $J_0(\dots)$ — функция Бесселя первого рода нулевого порядка от действительного аргумента. Вычисляя интеграл по ρ , находим

$$\begin{aligned} I &= 2\pi \int_0^r J_0(k\rho \sin\varphi) \rho d\rho = 2\pi r J_1(kr \sin\varphi) / (k \sin\varphi) = \\ &= 2\pi r^2 J_1(kr \sin\varphi) / (kr \sin\varphi). \end{aligned} \quad (1.6)$$

Мощность сигнала на выходе фотоприемника пропорциональна I^2 :

$$P \sim [J_1(kr \sin\varphi) / (kr \sin\varphi)]^2 = [J_1(2\pi r / \lambda \sin\varphi) / (2\pi r / \lambda \sin\varphi)]^2. \quad (1.7)$$

Из (1.7) следует, что для круглой апертуры мощность сигнала P определяется углом φ и не зависит от α . Для нее $W(\varphi) = [J_1((2\pi \times r / \lambda \sin\varphi) / (2\pi r / \lambda \sin\varphi))]^2$ при $r/\lambda = 500$ (рис. 1.12). При $\sin\varphi = 0,82\lambda/r$ и т.д.

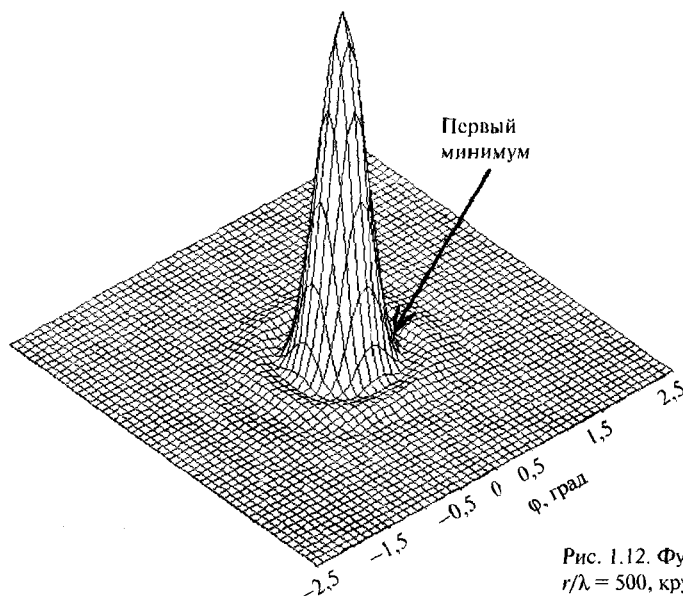


Рис. 1.12. Функция $W(\varphi)$ для $r/\lambda = 500$, круглая апертура

имеем максимумы; при $\sin \varphi = 0,61\lambda/r$, $1,12\lambda/r$ и т.д. — минимумы: $W(\varphi) = 0$. В оптике принято считать, что две точки разрешаются прибором, если угловое расстояние между ними $\Delta\varphi$ таково, что изображение первой точки попадает в основной максимум при $\varphi = 0$, а изображение второй — в первый минимум, отмеченный на рис. 1.12 стрелкой (*критерий Релея*). Следовательно, при малых значениях φ $\Delta\varphi = 0,61\lambda/r$, и пространственное разрешение $\Delta L = 0,61\lambda L/r$. В (1.7) не учтено затенение основного зеркала вспомогательным и некоторые другие малые эффекты.

Таким образом, чем больше отношение радиуса апертуры к длине волны, тем лучше разрешение оптической системы. Пусть, например, $\lambda = 1$ мкм, $r = 10$ см; $r/\lambda = 10^5$, тогда $\Delta\varphi = 6,1 \cdot 10^{-6}$ рад = $1,26''$. Если расстояние L от спутника до точки на поверхности равно 900 км, то пространственное разрешение $\Delta L = L \cdot \Delta\varphi = 5,5$ м. Из-за дифракции электромагнитных волн изображение точки несколько расплывается. Поэтому фотоприемник должен иметь конечный размер — порядка длины волны (размер так называемого второго кружка Эйри), в этом случае на фотоприемник попадает около 90% всей энергии. Таково предельное разрешение, определяемое размерами оптической системы и существенно зависящее от длины волны. При длине волны $\lambda = 10$ мкм и тех же условиях разрешение равно 55 м.

На практике представляет интерес изображение не точек, а достаточно протяженных предметов с размерами, превышающими предельное пространственное разрешение. В этом случае при $r/\lambda \gg 1$ изображение в фокальной плоскости можно построить по законам геометрической оптики. Справедливо выражение $d/D = F/L$, где D и d — размеры предмета и его изображения, F — фокусное расстояние объектива. Пусть $D = 1$ км, $F = 5$ м, $L = 900$ км. Тогда $d = 5,5$ мм. Установив в фокальной плоскости фотоприемники размером 5,5 мм, работающие в разных участках оптического диапазона, получим пространственное разрешение в 1 км в широком диапазоне длин волн. Оптическая система с главным зеркалом диаметром $2r = 20$ см установлена, например, в сканере AVHRR спутника NOAA с пространственным разрешением 1,1 км в подспутниковой точке для волн длиной от 0,56 до 12,5 мкм.

Для апертуры прямоугольной формы (на рис. 1.11 это может быть площадка $CDEF$) интегрирование в (1.5) проводится в прямоугольной системе координат x и y в пределах $-D_x/2 \leq x \leq D_x/2$, $-D_y/2 \leq y \leq D_y/2$, где $D_x = CE$, $D_y = CD$. Ток фотоприемника

$$I = \int_{-D_x/2}^{D_x/2} \int_{-D_y/2}^{D_y/2} \exp[-ik/L(xx_1 + yy_1)] dx dy. \quad (1.8)$$

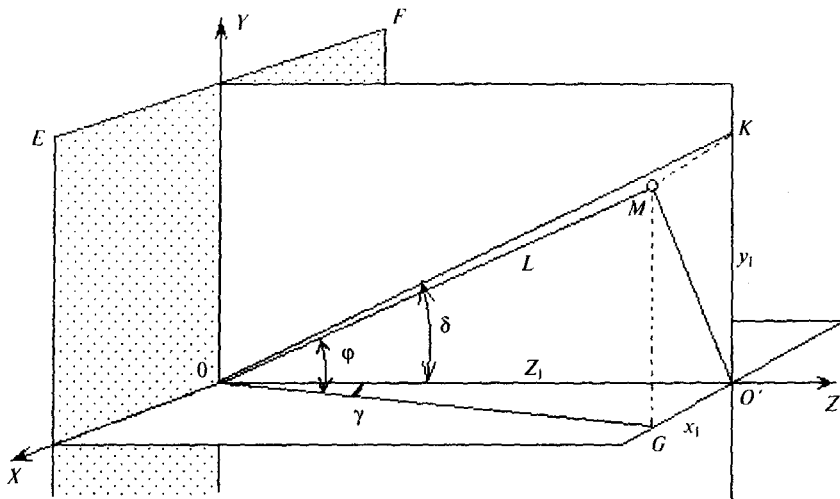


Рис. 1.13. Прямоугольная апертура

В этом случае вместо зависимости I от угла φ имеем зависимость от углов γ и δ , лежащих в плоскости XOZ и YOZ соответственно (рис. 1.13).

Если x_1, y_1, z_1 — по-прежнему координаты точки M , то $z_1 = OO' = L \cos \varphi$, $x_1 = MK = GO' = L \cos \varphi \cos \gamma$, $y_1 = O'K = MG = L \cos \varphi \sin \alpha$. Из этих соотношений нетрудно найти, что угол φ связан с углами γ и δ выражением $\cos^2 \varphi = 1 / (1 + \sin^2 \gamma + \sin^2 \delta)$. Интегрирование по x и y дает зависимость мощности на фотоприемнике от γ и δ :

$$W(\gamma, \delta) = \{(\sin \pi a D_x / \lambda) / \pi a D_x / \lambda\}^2 \cdot \{(\sin \pi b D_y / \lambda) / \pi b D_y / \lambda\}^2, \quad (1.9)$$

где $a = \sin \gamma / (1 + \sin^2 \gamma + \sin^2 \delta)^{1/2}$, $b = \sin \delta / (1 + \sin^2 \gamma + \sin^2 \delta)^{1/2}$.

Функция $W(\gamma, \delta)$ (рис. 1.14) имеет главный максимум при $\gamma = 0, \delta = 0$, минимумы — при $\pi a D_x / \lambda = \pi, 3\pi, \dots$; $\pi b D_y / \lambda = \pi, 3\pi, \dots$. Если размеры апертуры по x и y во много раз превышают длину волны, то мгновенное поле зрения по горизонтали $\Delta \gamma = \lambda / D_x$ и по вертикали $\Delta \delta = \lambda / D_y$.

Кроме оптического диапазона электромагнитных волн, дистанционное зондирование Земли ведется в радиодиапазоне, при этом вместо оптических систем на спутники устанавливаются антенны. И в этом случае полученные выше формулы для углового и пространственного разрешения полностью справедливы. Но длина волны в радиодиапазоне на несколько порядков больше длины волны в оптическом диапазоне. Например, спутник ERS на высоте $H = 780$ км принимает радиолокационные сигналы, отраженные от земной поверхности, на длине

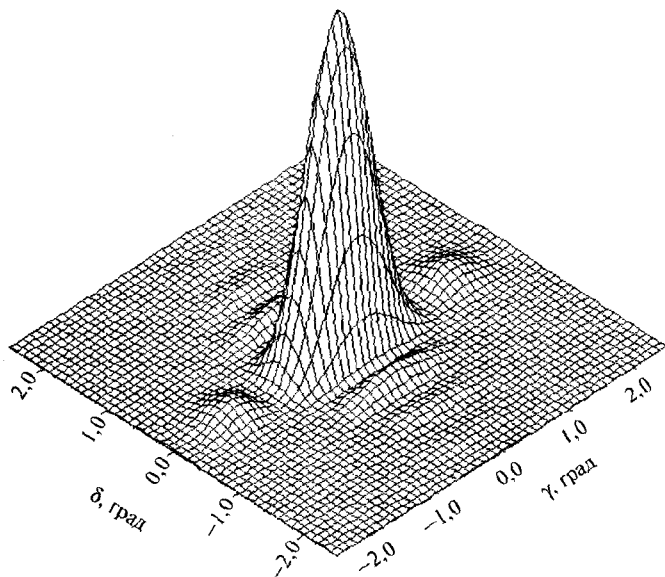


Рис. 1.14. Функция $W(\gamma, \delta)$ для $D_x/\lambda = 50$, $D_y/\lambda = 100$ (прямоугольная апертура)

Волны $\lambda = 5,3$ см при дальности $L = 900$ км. Нетрудно найти, что для достижения предельного разрешения $\Delta L = 30$ м диаметр приемной параболической антенны должен быть равен $2r = 2 \cdot 0,61\lambda L/\Delta L = 2$ км, что нереально. Однако известны способы получения в радиодиапазоне пространственного разрешения такого же порядка, что и в оптическом, при сравнительно небольших антеннах размером всего 8–15 м.

1.3.2. Радиотехнические методы

В общих чертах принцип активной радиолокации состоит в следующем. На спутнике устанавливается передатчик, посылающий с помощью антенны в направлении Земли импульсы с высокочастотным заполнением (рис. 1.15). После этого наступает пауза, в течение которой производится прием отраженных сигналов. Если импульс отражается от некоторого объекта M , расположенного на расстоянии L от спутника, то отраженный сигнал вернется назад через интервал времени $\Delta t = 2L/c$, где c — скорость света, множитель 2 учитывает, что сигнал проходит путь L дважды: от радиолокатора до объекта и от объекта до радиолокатора. Чем дальше объект от радиолокатора, тем больше Δt . Интенсивность отраженных сигналов зависит от дальности и различна

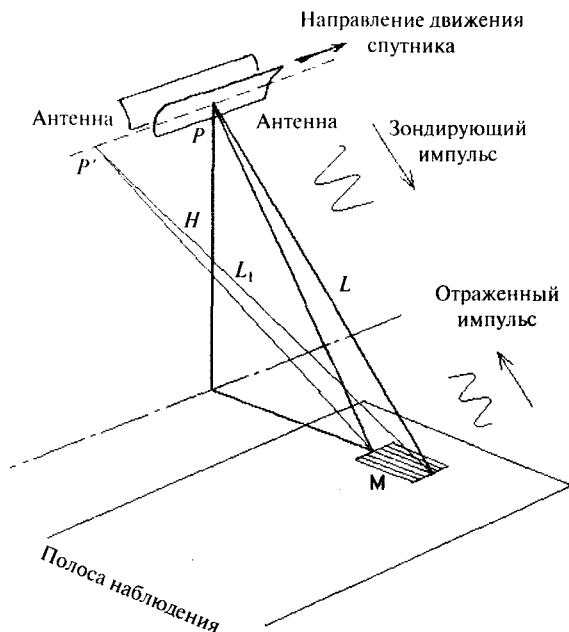


Рис. 1.15. Схема работы радиолокатора бокового обзора

для различных объектов, так как они отличаются размерами и электрофизическими характеристиками. Измеряя Δt , можно найти расстояние до объекта. Таким образом, средствами радиолокационной техники автоматически осуществляется сканирование по дальности, так как сигналы от разных объектов приходят в разное время.

Для достижения высокого пространственного разрешения вдоль строки необходимо использовать очень короткие им-

пульсы, поскольку электромагнитная волна распространяется со скоростью света, проходя 300 м за 1 мкс. Укорочение импульса приводит к уменьшению его энергии, что не всегда приемлемо, поэтому высокочастотное заполнение сравнительно длинного импульса (длительностью в несколько микросекунд) модулируют в передатчике специальным образом, а отраженный сигнал в приемнике сжимают (укорачивают). Для современной техники разрешение в 5–10 м не является пределом. Радиолокатор перемещается вместе со спутником, последовательно считывая по строкам сигнал интенсивностью, пропорциональной отражательной способности различных участков поверхности. Строки, как и в сканерах оптического диапазона, расположены поперек движения спутника. Отсюда следует, что антенна радиолокационной станции, принимающая отраженные сигналы, должна быть направлена именно в этом, боковом направлении (см. рис. 1.15), поэтому такого рода устройства называют радиолокационными станциями бокового обзора (РЛС БО).

Пространственная разрешающая способность РЛС БО в направлении движения спутника (разрешение между строками) зависит от направленных свойств приемной антенны. Антенна выполняет те же

функции, что и оптическая система на рис. 1.10, суммируя в пределах апертуры энергию, приходящую от некоторого участка местности M на поверхности. Чем меньше этот участок, тем лучше разрешение. Зависимость мощности на выходе антенны от углов γ и δ , называемая *диаграммой направленности* антенны по мощности, аналогична показанной на рис. 1.14.

Как правило, спутниковые антенны имеют прямоугольную апертуру, мгновенное поле зрения антенны по горизонтали $\Delta\gamma = \lambda/D_x$, по вертикали $\Delta\delta = \lambda/D_y$, где D_x , D_y — размеры апертуры по горизонтали и по вертикали. На советском спутнике «Алмаз-1А» были установлены две антенны с $D_x = 15$ м, $D_y = 1,5$ м, $\lambda = 10$ см для каждой, высота орбиты над землей $H = 280$ км. Мгновенное поле зрения по горизонтали составляло $\Delta\gamma = 0,38^\circ$, по вертикали $\Delta\delta = 3,8^\circ$; для объектов на расстоянии $L = 500$ км от спутника пространственное разрешение по горизонтали (вдоль направления движения спутника) $\Delta L = L\Delta\gamma = 3,3$ км. Предел увеличения реального (физического) размера апертуры бортовой антенны спутника близок к 15 м. К счастью, это ограничение можно обойти, используя радиолокаторы с синтезированной апертурой (РСА).

В принципе идея функционирования РСА достаточно проста: используется движение самого спутника и установленной на нем антенны с реальной апертурой (скорость более 7 км/с). Сигналы, принятые антенной, запоминаются и далее суммируются (накапливаются) компьютером в соответствии с выражением (1.8), так что сигнал от объекта M , принятый в точке P (см. рис. 1.15) в момент времени t_p , складывается с сигналом от этого же объекта, который ранее принимался в момент t_m , когда спутник находился в точке P' . Для объектов, находящихся ближе к спутнику, используют меньшую синтезированную апертуру, для дальних объектов размер апертуры растет пропорционально росту расстояния L . Это позволяет уменьшать мгновенное поле зрения $\Delta\gamma$ для дальних объектов и делать разрешение независимым от дальности, что недостижимо в оптическом диапазоне. При $\lambda = 5,3$ см для обеспечения пространственного разрешения в 30 м на дальности 1000 км необходима синтезированная апертура 1,77 км, на дальности $L = 1200$ км — 2,12 км.

На практике применяются как РЛС БО с реальной апертурой (их называют также некогерентными РЛС БО), так и РСА — так называемые когерентные РЛС БО. Преимуществом некогерентных РЛС БО являются более широкая полоса обзора и относительная простота как самого радиолокатора, так и системы обработки информации. Радиолокационные системы с синтезированной апертурой позволяют получать наибольшее высокое разрешение, но требуют сложной системы обработки



Рис. 1.16. Радиолокационное изображение горной местности на юге Красноярского края (разрешение 100 м, вертикальная поляризация)

на борту. В целом же пространственное разрешение РЛС БО (10–100 м для РСА и 1–2 км для некогерентных РЛС БО) сравнимо с разрешением оптических систем. На рис. 1.16 представлено радиолокационное изображение горной местности на юге Красноярского края с разрешением 100 м, полученное с помощью РСА, установленного на космическом пилотируемом аппарате Space Shuttle (США).

Для сигналов РЛС очень важна их чувствительность к содержанию воды в объектах, так как наличие воды увеличивает проводимость среды и интенсивность отражения от нее. Как и в оптическом диапазоне, в радиоволновом сигналы различной длины волны несут различную информацию об окружающей среде. В частности, для густой растительно-

сти интенсивность отражения в пределах сантиметрового диапазона растёт приблизительно обратно пропорционально длине волны, а для редкой растительности — обратно пропорционально ее квадрату.

Для работы в радиодиапазоне весьма важна *поляризация* отраженной волны — направление вектора напряженности электрического поля E . РЛС может излучать сигналы с горизонтальной поляризацией (вектор E расположен горизонтально) либо с вертикальной (вектор E расположен вертикально), а иногда применяют оба вида поляризации: горизонтальную на одной длине волны, вертикальную — на другой. Отраженная от объекта волна может частично менять свою поляризацию, поэтому приемная антенна спутника нередко построена так, чтобы принимать сигналы с двумя видами поляризации на каждой частоте. Сравнивая эти сигналы, т.е. оценивая анизотропию поляризации сигнала, можно получить дополнительные сведения об объекте, его структуре и электрофизических характеристиках. Если средства дистанционного зондирования оптического диапазона наиболее эффективны при изучении растительности, обнаружении пожаров, оценке температуры поверхности, то активные средства, работающие в радиодиапазоне, перспективны для получения сведений о почве и геологических структурах, при изучении водоемов, льдов на суше и на воде, в океанологии и несколько в меньшей степени для изучения растительности. Качество радиолокационной съемки не зависит от освещенности поверхности Земли и от наличия облачного покрова, что выгодно отличает эти системы от оптических средств дистанционного зондирования.

Космические платформы, оснащенные бортовыми радиолокаторами, — наиболее дорогостоящие, крупногабаритные и массивные спутники из всех аппаратов, предназначенных для исследования Земли. В этом смысле рекордсменом был спутник «Алмаз-1А» с когерентной РЛС БО, имевший массу 18,55 т. Отметим, что на спутники одновременно с РЛС БО, как правило, устанавливают и аппаратуру дистанционного зондирования оптического диапазона.

К средствам активного радиолокационного зондирования относят также высотомеры и скаттерометры. Радиолокационные высотомеры применяют для измерения высотного профиля подстилающей поверхности с точностью 2–8 см и для получения информации о форме морской поверхности, гравитационных аномалиях, высоте волн, скорости ветра, уровнях приливов, скорости поверхностных течений, ледовом покрове и т.д.

Принцип действия скаттерометров (измерителей характеристик рассеяния) основан на зависимости эффективной площади рассеяния морской поверхности и ее анизотропии от скорости и направления ветра. Основным назначением их является определение синопти-

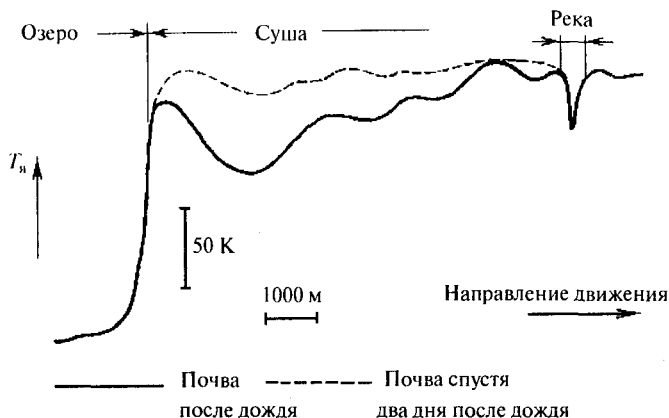


Рис. 1.17. Радиояркостные температуры участка поверхности, снятые с высоты 50 м ($\lambda = 20$ см, вертикальная поляризация)

ческого поля ветра, что не требует высокого пространственного разрешения; скаттерометры создаются на основе РЛС с непрерывным излучением.

В заключение кратко остановимся на *пассивном* радиотехническом методе наблюдения земной поверхности из космоса — радиометрическом зондировании в микроволновом диапазоне (частоты 1–100 ГГц). Как и приборы дальнего ИК-диапазона, радиометры регистрируют собственное тепловое излучение поверхности (рис. 1.17). Обычно их калибруют в радиационных (радиояркостных) температурах T_y . По сравнению с зондированием в ИК-области спектра радиометрический метод обладает важными преимуществами: возможностью получения информации о параметрах верхнего слоя грунта (например, о влажности на глубинах до 1–2 м), о параметрах ледяного покрова, морского волнения и др. В этом диапазоне волн атмосфера практически прозрачна. По сравнению с ИК в радиодиапазоне наблюдаются значительные яркостные контрасты при одинаковых температурах объектов.

Вместе с тем радиометрическим методам свойственны и принципиальные недостатки: меньшее угловое разрешение, чем при инфракрасной радиометрии, а также более низкая абсолютная точность измерения температуры, так как в соответствии с формулой Планка при обычной температуре плотность потока мощности излучения в ИК-диапазоне во много раз больше, чем в микроволновом.

1.3.3. Спутниковые методы исследования атмосферы

Искусственные спутники Земли позволяют не только наблюдать из космоса поверхность суши, водоемов и облаков, но и определять средними оптической спектроскопии концентрацию некоторых газов и аэрозолей.

Естественные и антропогенные примеси, вызывающие локальное загрязнение территорий, могут разноситься потоками воздуха по всему земному шару. Например, выбросы Норильского горно-металлургического комбината заметны на Аляске и в Канаде, в Японии идут кислотные дожди из-за промышленных выбросов в Китае. Основная роль в выявлении глобального загрязнения атмосферы отводится спутниковым методам. Для оценивания содержания малых газов, CO_2 и аэрозолей используют спутниковые спектрофотометры. На рис. 1.18, построенном по данным спутника TOMS/EP за 1 октября 1994 г., видны выбросы SO_2 при извержении вулкана Ключевская сопка (отмечена крестом), Норильского комбината (стрелка) и выбросы из Китая (внизу рисунка).

Спектрофотометры УФ- и видимого диапазонов регистрируют интенсивность рассеянного «назад» излучения Солнца. Спектрофотометры ИК-диапазона фиксируют интенсивность прошедшего через атмо-

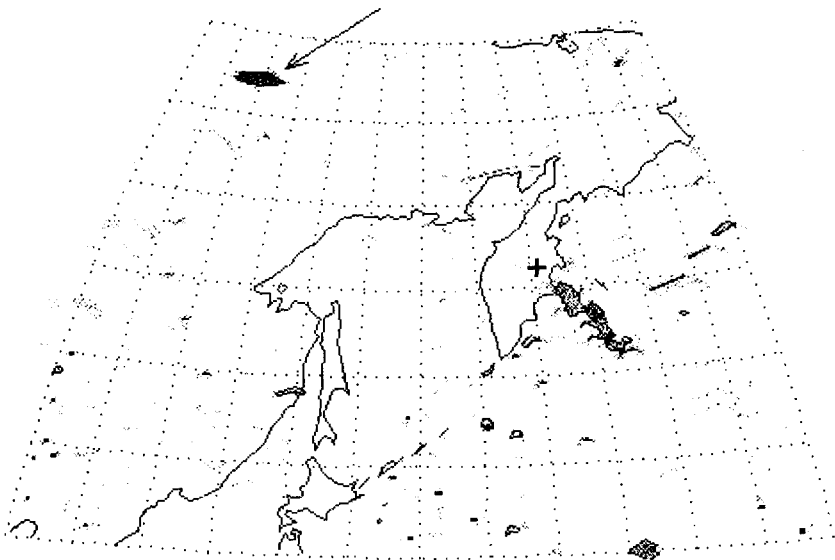


Рис. 1.18. Карта выбросов SO_2 по спутниковым данным за 1 октября 1994 г.

сферу теплового излучения от поверхности Земли и облаков. Частицы аэрозолей, как правило, имеющие несферическую форму, под действием воздушных потоков ориентируются приблизительно в одном направлении, поэтому солнечный свет, рассеянный аэрозолями, имеет эллиптическую поляризацию. Измеряя характеристики поляризации рассеянного излучения, можно оценить концентрацию аэрозолей.

При определении спутниковыми методами общего содержания озона O_3 (ОСО) в атмосфере используются интенсивные полосы поглощения озона в УФ- и ИК-областях.

1.4. ДИСТАНЦИОННОЕ ЗОНДИРОВАНИЕ ЗЕМЛИ ИЗ КОСМОСА

1.4.1. Орбиты спутников

Траектория движения искусственного спутника Земли называется его орбитой. При выключенных маршевых реактивных двигателях свободное движение спутника под действием гравитационных сил и по инерции подчиняется законам небесной механики. Считая Землю строго сферической с равномерным распределением массы внутри нее, а действие гравитационного поля Земли единственной силой, действующей на спутник, можно решить так называемую *задачу Кеплера*, которая сводится к уравнению кривой второго порядка — эллипса (или окружности — частного случая эллипса):

$$m d^2 \mathbf{r} / dt^2 = -\gamma m M \mathbf{r} / r^3, \quad (1.10)$$

где m — масса спутника, $M = 5,976 \cdot 10^{27}$ г — масса Земли, \mathbf{r} — радиус-вектор, соединяющий спутник и центр Земли, r — его модуль, $\gamma = 6,67 \cdot 10^{-14}$ м³/гс³ — гравитационная постоянная. Решая уравнение в полярных координатах r, ν , получаем

$$r = p / (1 + e \cos \nu). \quad (1.11)$$

Эллиптическая орбита, по которой вращается спутник (рис. 1.19, где в точке S находится спутник, а в точке G — Земля), характеризуется следующими параметрами: $a = AO$ и $b = OC$ — *большая и малая полуоси* эллипса; $e = (1 - b^2/a^2)^{1/2}$ — *эксцентриситет орбиты*; угол PGS — *угловая координата ν радиуса-вектора* (так называемая *истинная аномалия*); *фокальный параметр* $p = b^2/a$; $p = K^2/\gamma m^2 M$, где K — момент количества движения спутника. К параметрам орбиты спутника относится также

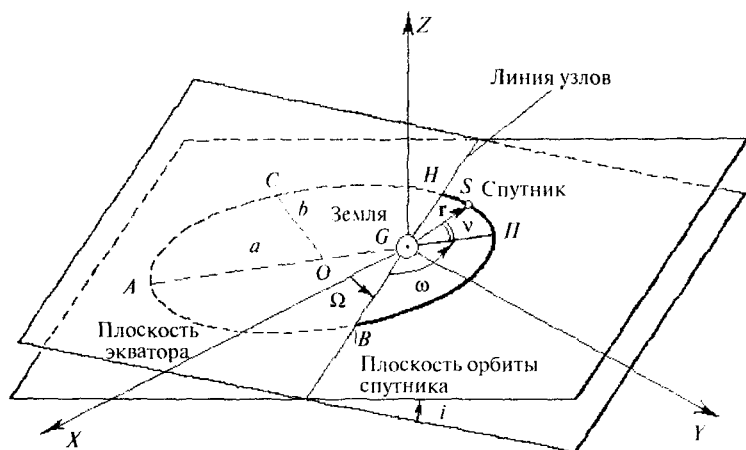


Рис. 1.19. Эллиптическая орбита

период обращения T — время между двумя последовательными прохождением одной и той же точки орбиты.

В рамках задачи Кеплера спутник движется в *плоскости орбиты*, проходящей через центр Земли. В так называемой *абсолютной* или *звездной* системе координат плоскость орбиты неподвижна. Абсолютная система — это декартова система координат с началом в центре Земли, неподвижная относительно звезд. Ось Z направлена вдоль оси вращения Земли и указывает на север, ось X направлена на точку весеннего равноденствия, в которой находится Солнце 21 марта в 0 ч по всемирному времени, а ось Y перпендикулярна осям X и Z .

В общем случае плоскость орбиты пересекается с плоскостью экватора Земли по так называемой *линии узлов* (см. рис. 1.19). Точка B , в которой орбита пересекает плоскость экватора при движении спутника с юга на север, называется *восходящим узлом* орбиты, точка H пересечения при движении спутника с севера на юг — *нисходящим узлом*. Положение восходящего узла определяется *долготой восходящего узла*, т.е. углом Ω между восходящим узлом и точкой весеннего равноденствия, отсчитываемым против часовой стрелки, если смотреть со стороны Северного полюса. Для линии узлов задают два угла в плоскости орбиты. Угол ω — угловое расстояние, отсчитываемое от восходящего узла в плоскости орбиты до *перигея орбиты* P , т.е. ближайшей к Земле точки орбиты спутника; ω называют *аргументом перигея*. Угол i между плоскостью орбиты и плоскостью экватора, называемый *наклоением орбиты*, отсчитывается от плоскости экватора с восточной стороны восходящего узла орбиты, против движения часовой стрелки. По накло-

нению различают *экваториальные* ($i = 0^\circ$), *полярные* ($i = 90^\circ$) и *наклонные* ($0 < i < 90^\circ$, $90 < i < 180^\circ$) орбиты.

Долгота восходящего узла Ω , наклонение i и аргумент перигея ω характеризуют положение плоскости орбиты и ее ориентацию в пространстве. Форму и размер орбиты задают фокальный параметр p и эксцентриситет e . Для привязки движения спутника ко времени в число элементов вводится время прохождения спутником точки начала отсчета t_0 . Совокупность параметров Ω , ω , i , p , e , t_0 называется *кеплеровскими элементами* или *элементами орбиты*.

Зная параметры Ω , ω , i , p , e и положение спутника на орбите в момент t_0 , можно найти это положение в любой другой момент времени t_1 .

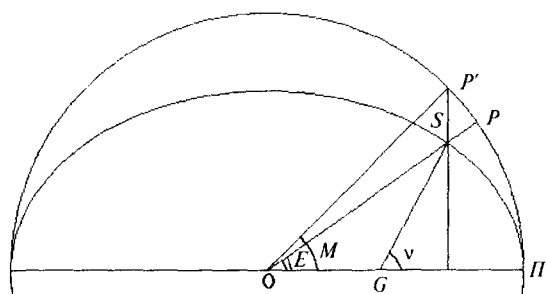


Рис. 1.20. Диаграмма, иллюстрирующая вычисление положения спутника

Пусть спутник движется вокруг Земли G по эллиптической орбите. Проведем из центра этой орбиты O окружность радиусом, равным большой полуоси эллипса (рис. 1.20). Предположим, что в момент t_0 спутник находился в перигелии орбиты P , а в момент t_0 сместился в точку S .

Угол $\angle PGS$ (между направлением на перигелий и радиусом-вектором), как указывалось, называется *истинной аномалией* ν в момент t_0 . Проведем через S прямую, перпендикулярную к оси OP и пересекающуюся в точке P' с окружностью. Угол $\angle POR$ называется *эксцентрической аномалией* E в момент t_0 . Представим теперь точку, которая выходит из перигелия одновременно со спутником и движется по окружности равномерно со скоростью, равной средней скорости движения спутника по орбите. Эта средняя скорость называется *средним движением* и равна $n = 360^\circ/T$, где T — период обращения. Если в момент t_0 такая точка займет положение P' , то угол $\angle POP'$ будет равен $M = n(t_0 - t_0)$. Эта величина называется *средней аномалией* в момент t_0 . Решая трансцендентное уравнение

$$E - e \sin E = M,$$

называемое уравнением Кеплера, можно найти эксцентрическую аномалию E . Истинная аномалия ν , характеризующая положение спутни-

на орбите в абсолютной системе координат в момент t_0 , связана с E и эксцентриситетом e соотношением

$$\operatorname{tg} v/2 = [(1 + e)/(1 - e)]^{1/2} \operatorname{tg} E/2.$$

Зная среднее движение n и истинную аномалию v в момент t_0 , можно вычислить t_n и далее истинную аномалию v в момент t_1 , т.е. определить положение спутника на орбите.

Однако кеплеровские элементы дают лишь приближенное описание орбиты спутника. Во-первых, массы внутри Земли распределены неравномерно. Во-вторых, на движение спутника влияет сопротивление земной атмосферы. В-третьих, необходим учет светового давления солнечных лучей. В-четвертых, нужно учитывать притяжение Луны и Солнца и др. Влияние этих сил на движение ИСЗ мало по сравнению с силой притяжения Земли. Они называются *возмущающими силами*, а движение спутника с учетом их воздействия — *возмущенным движением*. Основным источником возмущений является первый фактор. Если учитывать только первую зональную гармонику в разложении гравитационного потенциала Земли (она описывает сжатие Земли с полюсов), то окажется, что в основном изменяется ориентация орбиты в пространстве, а форма и размеры орбиты остаются постоянными. За один оборот долгота восходящего узла Ω и аргумент перигея ω изменяются на

$$\Delta\Omega = -0^\circ,58 (R_0/a)^2 \cos^2 i / (1 - e^2)^2,$$

$$\Delta\omega = 0^\circ,29 (R_0/a)^2 (5\cos^2 i - 1) / (1 - e^2)^2,$$

где $R_0 = 6378,14$ км — экваториальный радиус. Эти выражения, в первом приближении определяющие поправки к долготе восходящего узла Ω и аргументу перигея ω , позволяют уточнить положение орбиты в абсолютной системе координат.

Спутник, движущийся в земной атмосфере, испытывает аэродинамическое торможение, зависящее от плотности атмосферы на высоте полета, от скорости спутника, площади его поперечного сечения и массы. Возмущение орбиты за счет аэродинамического торможения содержит регулярную и нерегулярную составляющие. К регулярным возмущениям приводит суточный эффект (ночью, т.е. в конусе земной тени, плотность атмосферы на данной высоте меньше, чем днем). Движение воздушных масс, влияние потоков заряженных частиц, выбрасываемых Солнцем, приводят к нерегулярным возмущениям. Для природоведческих спутников сопротивление атмосферы играет заметную роль только при низких орбитах; при высоте перигея более 500—600 км возмуща-

ющее ускорение от неравномерности распределения масс превышает на два порядка и более ускорение от торможения в атмосфере.

При высоте перигея от 500–600 до нескольких тысяч километров к основному возмущающему фактору добавляется давление солнечного света (вместо сопротивления атмосферы). Влияние этого давления проявляется в дополнительных малых периодических возмущениях элементов орбиты. Если же спутник движется так, что регулярно попадает в конус земной тени, то имеют место также и небольшие постоянные изменения элементов. Но ускорение за счет давления света на несколько порядков меньше возмущающего ускорения за счет основного фактора. Еще слабее влияние притяжения Луны и Солнца.

Зная орбитальные элементы, можно предсказать время прохождения спутника над тем или иным районом, направить антенну приемной станции на спутник, выполнить географическую привязку изображений, полученных с помощью спутника. Существуют модели движения спутников по орбите, учитывающие все основные возмущающие силы. Они позволяют с высокой точностью предсказывать значения орбитальных элементов на многие месяцы вперед. Примером является модель, разработанная в Институте космических исследований РАН. Однако в практике дистанционного зондирования чаще всего применяют орбитальные элементы в формате NORAD (США) из сети Интернет (<http://www.celestrak.com>). По этому адресу представлены орбитальные элементы для спутников самого различного назначения, запущенных в разное время в различных странах, в том числе архивные материалы. Формат NORAD использует достаточно простые модели движения спутников, но орбитальные элементы постоянно корректируются по данным сети радиолокационных станций и станций оптического наблюдения (для низколетящих спутников корректировка может производиться несколько раз в день). Достоинством элементов в формате NORAD является их доступность, однако эта информация сравнительно быстро устаревает. В частности, если при расчете орбиты спутника NOAA использовать элементы NORAD недельной давности, то окажется, что подспутниковая точка сместится примерно на 50 км относительно ее истинного положения.

Формат NORAD содержит две строки данных, каждая по 69 символов. Рассмотрим пример:

NOAA 14

```
1 23455U 94089A 99244.00000000 .00000328 00000-0 20432-3 0 243
2 23455 99.1001 212.1966 0008831 180.9531 7.7697 14.12021817 2 40644
```

Цифры 1 и 2 в начале каждой строки — это номер строки. Далее идет классификационный номер спутника по каталогу NORAD, для NOAA-14 — это 23455. Буква U после номера спутника в первой строке указывает, что данные общедоступны.

В первой строке: 94089А — номер спутника по международной классификации, причем 94 — последние цифры года запуска (1994 г.), 089А — номер запуска в 1994 г.

В следующем поле приведено время $t_0 = 99244,00000000$, к которому относятся орбитальные элементы. Первые две цифры 99 указывают на год, т.е. 1999; 2000 г., таким образом, соответствуют цифры 00; 2001 г. — 01; 244 — это номер дня с начала года, отсчитываемый с 00 ч 1 января по Гринвичу, т.е. 1 сентября 1999 г. Цифры после точки 00000000 указывают на $1/100\,000\,000$ долю суток, в данном случае это 00 ч 1 сентября 1999 г. по Гринвичу.

Следующие два поля: .00000328 и 00000-0 — это первая производная среднего движения, деленная на 2 (в единицах оборота за день в квадрате), и вторая производная, деленная на 6 (в единицах оборота за день в кубе). В орбитальной модели формата NORAD эти сведения не используются.

Величина 20432-3 — это сокращенная форма записи: $B^* = 0,20432 \cdot 10^{-3}$. В орбитальной модели NORAD B^* характеризует аэродинамическое торможение данного спутника при движении его в атмосфере на высоте орбиты. Далее представлен номер орбитальной модели. В формате NORAD это всегда модель SGP4/SDP4, на что указывает цифра 0.

Цифра 24 в принципе должна отражать номер сеанса коррекции орбитальных данных, но обычно это достаточно произвольная величина. Последняя цифра 3 в первой строке, как и 4 во второй, — контрольная сумма — остаток от деления на 10 суммы всех цифр в строке, причем буквы, пробелы, десятичные точки, знаки «+» игнорируются, а знаком «-» присваивается 1.

Во второй строке повторяется классификационный номер спутника по каталогу NORAD, далее приводятся орбитальные элементы для момента t_0 : $i = 99,1001$ — угол наклона орбиты (град); $\Omega = 212,1966$ — долгота восходящего узла в градусах. Следующее число 0008831 — это сокращенное обозначение эксцентриситета орбиты, который равен в данном случае $e = 0,0008831$. Число $\omega = 180,9531$ — аргумент перигея в градусах; $M = 7,7697$ — средняя аномалия в градусах; $n = 14,12021817$ — среднее движение в оборотах за сутки, связанное с периодом обращения спутника T соотношением $n = 1440/T$, здесь T выражено в мин; 240664 — номер витка с момента запуска спутника. Последнее число 4, как указывалось, это контрольная сумма.

Чтобы определить время прохождения спутника над тем или иным районом Земли, найти углы поворота антенны приемной станции в направлении на спутник, необходимо перейти от абсолютной системы координат в систему, связанную с Землей и учитывающую ее движение. Соответствующие программные продукты, ориентированные на формат NORAD, достаточно хорошо представлены в Интернет (списки программных продуктов можно найти, например, по указанному выше адресу <http://www.celestrak.com>).

Спутники для дистанционного зондирования Земли запускают в основном на круговые орбиты. Малое значение эксцентриситета орбиты спутника NOAA-14, равное $e = 0,0008831$, достаточно типично. Такой спутник пролетает над различными участками Земли на одинаковой высоте, что обеспечивает равенство условий съемки. В этом случае справедливо соотношение

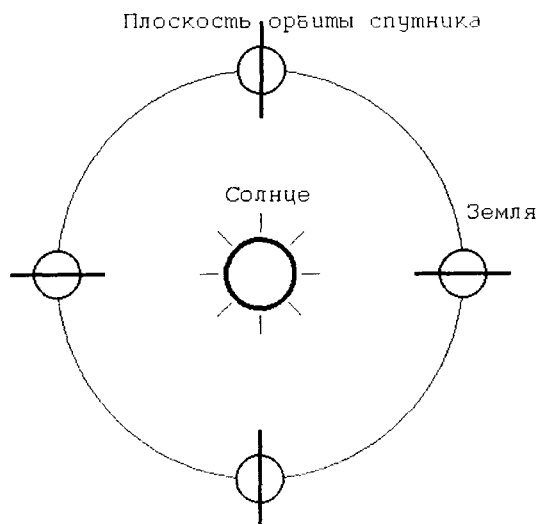
$$mV^2/R = mMy/R^2.$$

В левой части стоит центробежная сила, справа — сила притяжения спутника к Земле. Здесь m — масса спутника, V — скорость его на орбите, $M = 5,976 \cdot 10^{27}$ г — масса Земли, $R = R_0 + H$ — расстояние между спутником и центром Земли, причем $R_0 = 6370$ км — радиус Земли, H — высота спутника над поверхностью Земли, γ — гравитационная постоянная. Таким образом, $V = (M\gamma/R)^{1/2}$, период обращения спутника $T = 2\pi R/V$. Обозначим: $B = (M\gamma)^{1/2} = 6,31 \cdot 10^2$ км^{3/2}/с. Тогда $V = B/R^{1/2}$, $T = 2\pi R^{3/2}/B$.

Скорость перемещения подспутниковой точки по поверхности Земли V_3 может быть определена по формуле $V_3 = VR_0/R$.

Пусть $H = 1000$ км, тогда $R = 7370$ км. Используя приведенные формулы, находим, что скорость на орбите $V = 7,35$ км/с, $V_3 = 6,35$ км/с, период обращения $T = 105$ мин.

Низкоорбитальные спутники ($H < 1000$ км) обычно выводятся на приполярные солнечно-синхронные орбиты. Эти орбиты имеют наклонение относительно экватора, близкое к 90° , обеспечивают съемку всей поверхности Земли, включая полярные области. Поворот орбиты относительно Земли синхронизован с вращением Земли относительно Солнца, так что в течение всего времени угол между плоскостью орбиты и направлением на Солнце постоянен (рис. 1.21). Это позволяет производить съемку приблизительно в один и тот же час местного времени в течение всего года. Наиболее удобное время для съемки — около 12 ч местного времени.



Спутник NOAA-12 имеет орбиту с наклонением $98,86^\circ$, так что за сутки плоскость орбиты поворачивается относительно Земли на $0,986^\circ$, за год — на 360° . Однако из-за возмущений, вызванных неравномерностью распределения масс Земли, сопротивлением атмосферы, давлением солнечных лучей и т.п., за несколько лет плоскость орбиты меняется.

Ряд метеорологических спутников запуска-

Рис. 1.21. Солнечно-синхронная орбита

ются на геостационарную орбиту высотой 36 тыс. км с периодом 24 ч. Эти спутники находятся над одной и той же точкой на экваторе и обеспечивают постоянный обзор одной и той же части планеты от 50° с. ш. до 50° ю. ш.

1.4.2. Прием спутниковой информации

Станции для приема информации со спутников на Земле (называемые *земными*) содержат антенну с опорно-поворотным устройством (ОПУ), радиоприемное устройство и средства обработки, хранения и отображения информации (рис. 1.22).

Наиболее употребительные зеркальные антенны с параболическим рефлектором наводятся ОПУ на спутник по командам компьютера, в который заложены орбитальные данные. В фокусе антенны установлен облучатель, сигнал с которого усиливается маломощным усилителем (МШУ). Далее сигнал по кабелю поступает на приемник, цифровой сигнал с выхода которого обрабатывается на компьютере.

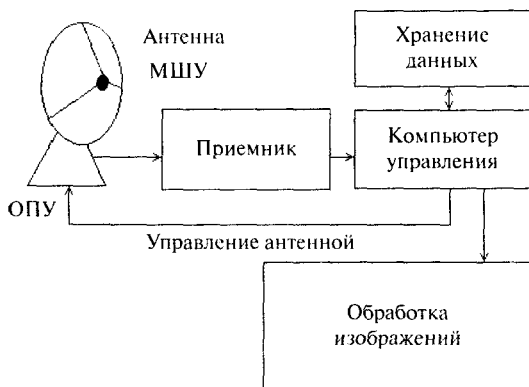


Рис. 1.22. Станция для приема информации с природо-ведческих спутников

Наиболее дорогостоящей частью станции является антенна с ОПУ. Чаще всего используются ОПУ с азимутально-угломестной подвеской антенны, позволяющие разворачивать ее на $\pm 180^\circ$ по горизонтали и на $0-90^\circ$ по углу места, отсчитываемому от горизонта к зениту. Азимутально-угломестная подвеска обладает принципиальным недостатком: в области углов места, примыкающих к зениту, образуется «мертвая зона», в пределах которой невозможно обеспечить связь со спутником. Это объясняется тем, что с ростом угла места Ψ требуемая угловая скорость вращения антенны вокруг вертикальной оси возрастает, стремясь к бесконечности при $\Psi \rightarrow 90^\circ$. Поскольку реальная скорость поворота антенны конечная, то, начиная с некоторого значения угла места, луч антенны отстанет от перемещения спутника, и сопровождение срывается. Таким образом, когда спутник близок к зениту, такой вид подвески не позволяет качественно принимать изображения той местности, где находится станция.

Для устранения «мертвой зоны» при прохождении спутника через зенит можно ввести в ОПУ третью ось. Однако в этом случае конструкция ОПУ резко усложнится. Во избежание этого можно сохранить двухосное поворотное устройство, но разместить ортогональные оси так, чтобы «мертвая зона» находилась в наименее существенной для поддержания связи части небесной полусферы, например ближе к горизонту.

При выборе конструкции антенны приходится учитывать различные факторы, в частности особенности распространения радиоволн на трассе Земля — космос. Для передачи сигналов с природоведческих спутников чаще всего используют радиоволны дециметрового и сантиметрового диапазонов или соответственно частоты 300 МГц—30 ГГц. В этом частотном диапазоне отдельные полосы переуплотнены различными радиослужбами. Так, полоса 300 МГц—10 ГГц интенсивно используется наземными радиостанциями. При этом повышается уровень взаимных помех, снижается качество радиосвязи.

При прохождении радиоволн сквозь атмосферу Земли приходится учитывать влияние тропосферы (0—11 км) и ионосферы (выше 80 км), поскольку в указанном интервале частот они несколько затухают в атмосферных газах и осадках. При этом изменяется поляризация волны, возникают дисперсионные искажения.

При прохождении через ионосферу линейно-поляризованные (в частности, горизонтально и вертикально поляризованные) радиоволны расщепляются на два эллиптически поляризованных компонента (обыкновенный и необыкновенный), которые распространяются с разной скоростью из-за влияния магнитного поля Земли. В результате сложения этих компонентов в точке приема плоскость поляризации результирующей волны будет повернута на некоторый угол (*эффект Фарадея*), зависящий от электронной концентрации N_e в ионосфере и напряженности геомагнитного поля H вдоль пути радиоволн в ионосфере. Для него характерна регулярная зависимость от времени суток, сезона и фазы цикла солнечной активности, а также случайные изменения, связанные с геомагнитными бурями и нерегулярными ионосферными неоднородностями. На частоте 1 ГГц угол поворота лежит в пределах $1\text{--}100^\circ$ и уменьшается с ростом частоты как $1/f^2$. Эффект поворота плоскости поляризации учтен в конструкции антенны: выбираются антенны и облучатели, способные принимать сигналы с круговой поляризацией, например спиральные антенны и спиральные облучатели.

При прохождении через ионосферу широкополосные сигналы искажаются, поскольку время распространения составляющих его спектра будет различно. Это явление, известное как относительная дисперсия, характеризуется разностью задержек между нижней и верхней частотами спектров сигналов, распространяющихся через ионосферу.

Относительная дисперсия зависит от N_c и H и обратно пропорциональна f^3 ; на частоте 1 ГГц может иногда достигать 0,4 нс/МГц и приводить к искажению сигналов, при полосе частот 100 МГц это 0,4 мкс.

Мощность сигнала в месте приема может быть оценена из следующих соображений. Если L — расстояние между передатчиком и приемником, $P_{\text{пер}}$ — мощность передатчика, то при условии, что излучение энергии происходит равномерно по всем направлениям (изотропный излучатель), вся энергия распределяется по площади сферы радиусом L , равной $4\pi L^2$. Мощность, приходящаяся на 1 м^2 , т.е. плотность потока мощности,

$$P = P_{\text{пер}}/4\pi L^2.$$

Реально спутник передает информацию только в нижнюю полушару, в сторону Земли. Поэтому приведенное выражение следует умножить на так называемый коэффициент направленного действия антенны (КНД) $D \geq 1$ — отношение плотности потока мощности, излучаемой антенной в направлении максимума ее диаграммы направленности (см. рис. 1.11 и 1.13), к плотности потока мощности, которая излучалась бы изотропным излучателем, при условии равенства общей излучаемой мощности. КНД связан с площадью апертуры S и длиной волны λ соотношением $D = 4\pi S/\lambda^2$. Если излучение происходит равномерно во всех направлениях в нижнюю полусферу, то $D = 2$. На природоведческих спутниках обычно устанавливают передающие антенны с $D = 3-4$, что позволяет земным станциям принимать информацию практически с любых направлений — от горизонта до горизонта. Таким образом,

$$P = P_{\text{пер}} D/4\pi L^2. \quad (1.12)$$

Приемная антенна — это барьер, поглощающий поток энергии, и излучаемый передающей антенной. Пусть площадь апертуры приемной антенны равна S . Если пренебречь потерями в приемной антенне, то мощность сигнала на ее выходе

$$P_{\text{пр}} = SP = SP_{\text{пер}} D/4\pi L^2. \quad (1.13)$$

В это выражение в явном виде не входит КНД приемной антенны, но с ростом S увеличивается отношение S/λ^2 , увеличивается D и сужается диаграмма направленности. В результате снижается уровень помех и шумов, которые могут поступать в антенну с боковых направлений. Однако слишком узкая диаграмма направленности требует большой точности наведения антенны.

Пусть радиус апертуры приемной параболической антенны $r = 60$ см; $P_{\text{пер}} = 5,5$ Вт; $D = 3$; $870 \text{ км} < L < 3400 \text{ км}$. Площадь апертуры антенны $S = \pi r^2 = 1,13 \text{ м}^2$, при $\lambda = 17,6$ см ее КНД около 400, ширина диаграммы направленности по ее первому минимуму, определяемая согласно (1.7) как $0,61\lambda/r$, около 10° . Эти реальные числа соответствуют мощности передатчика спутника NOAA, минимальному и максимальному расстоянию L от спутника до приемной станции, размеру антенны станции НРРТ для приема информации с этого спутника. Расчет по формуле (1.13) дает максимальное значение $P_{\text{пр}} = 2 \cdot 10^{-12}$ Вт, минимальное значение $P_{\text{пр}} = 10^{-13}$ Вт. Современная радиотехника позволяет усиливать и более слабые сигналы, но при этом усиливаются также внешние помехи и шумы и внутренние шумы радиоустройств.

Источниками внешних шумов в микроволновом диапазоне могут быть различные наземные радиопередатчики, существуют шумы и космического происхождения. Источником внутренних шумов радиоустройств прежде всего является дискретная природа электричества, так как электрический ток — это поток дискретных частиц — электронов.

Интенсивность шума принято описывать следующим образом. Все источники внешних и внутренних шумов заменяются эквивалентным источником шума в виде некоторого активного сопротивления (резистора). Известно, что на зажимах резисторов из-за хаотического теплового движения электронов возникает разность потенциалов, изменяющаяся случайным образом. Средняя мощность такого шума (его называют *тепловым*) описывается формулой Найквиста: $P = 4kT\Delta f$, где $k = 1,38 \cdot 10^{-23}$ Дж/град — постоянная Больцмана, T — температура резистора, Δf — полоса частот, в пределах которой измеряется средняя мощность шума. Если входное сопротивление приемника равно входному сопротивлению антенны (т.е. приемник и антенна согласованы), то эквивалентная мощность шума

$$P_{\text{ш}} = kT_{\text{ш}}\Delta f.$$

В нашем случае Δf — ширина полосы пропускания приемника, равная, в свою очередь, ширине полосы частот, необходимой для передачи информации со спутника, $T_{\text{ш}}$ — эквивалентная шумовая температура антенны и приемника, не совпадающая с термодинамической температурой, при которой находятся антенна и приемник. На прием сигналов с природоведческих спутников сильнее всего влияют внутренние шумы, и в первую очередь шумы первых каскадов усилителя радиосигналов. Поэтому во входных каскадах применяют малошумящие усилители (МШУ), которые конструктивно обычно совмещают с преобразовате

лем несущей частоты сигнала в более низкую и помещают непосредственно в облучателе антенны. Современные МШУ имеют в микроволновом диапазоне $T_{ш}$ порядка 40–70 К.

Пусть $T_{ш} = 70$ К, $\Delta f = 2$ МГц, что соответствует условиям приема сигналов со спутника NOAA. В этом случае $P_{ш} = 2 \cdot 10^{-15}$ Вт, что на 2–3 порядка меньше мощности сигнала.

Мощность сигнала при прочих равных условиях определяется размерами антенны и ее КНД, средняя мощность шума — шумовой температурой. Отношение мощности сигнала к средней мощности шума (*отношение сигнал–шум*) является важнейшей характеристикой качества приема и зависит, таким образом, от отношения КНД антенны к шумовой температуре. Эту величину $D/T_{ш}$ называют коэффициентом качества антенны. В рассмотренном примере коэффициент качества равен 5,7.

Выбор размеров приемной антенны определяется требованиями к коэффициенту качества и в конечном итоге — шириной полосы частот, необходимой для передачи информации со спутника. Последняя зависит от скорости передачи информации C . Для вычисления C необходимо знать параметры сканирующего устройства и скорость перемещения подспутниковой точки V_3 по Земле. Если разрешение сканера вдоль направления движения спутника равно ΔL , то в секунду считывается информация с $V_3/\Delta L$ строк. Пусть I — число бит, которое используется для записи яркости каждого пикселя, n — число спектральных каналов, K — коэффициент, зависящий от типа применяемого при передаче информации помехоустойчивого кодирования, $K > 1$, N — число пикселей в строке, связанное с шириной полосы обзора G соотношением $N = G/\Delta L$. Тогда

$$C = V_3 N I K n / \Delta L = V_3 G I K n / \Delta L^2. \quad (1.14)$$

Например, для $\Delta L = 1,1$ км, $V_3 = 6,56$ км/с, $G = 1670$ км, $I = 10$ бит, $n = 5$, $K = 1$ скорость передачи информации $C = 500$ кбит/с. Если $\Delta L = 100$ м, что было бы очень желательно, то при тех же условиях $C = 50$ Мбит/с. Улучшение пространственного разрешения приводит к увеличению информационного потока, который обратно пропорционален квадрату разрешения.

Полоса частот Δf , необходимая для передачи информации со спутника, зависит от вида модуляции высокочастотного колебания и ориентировочно равна $(3–3,5)C$. Для первого примера $\Delta f = 1,5$ МГц, для второго $\Delta f \geq 150$ МГц. Очевидно, что при прочих равных условиях средняя мощность шума для второго примера на два порядка выше. Чтобы со-

хранить необходимое отношение сигнал—шум, требуется увеличить площадь антенны и ее КНД в 100 раз, а диаметр антенны — в 10 раз. Таким образом, если при скорости передачи в 500 Кбит/с, пространственном разрешении 1,1 км и полосе обзора 1670 км можно применять антенну диаметром 1 м, то при скорости передачи 55 Мбит/с, пространственном разрешении 100 м с сохранением той же полосы обзора — антенну диаметром 10 м.

Типичная земная станция HRPT для приема информации со спутников NOAA имеет параболическую антенну диаметром 1,2—1,5 м. В фокусе антенны установлен облучатель, сигнал с которого усиливается МШУ, а несущая частота сигнала преобразуется в более низкую. МШУ имеет $T_{ш} = 60\text{--}80$ К. Далее сигнал по кабелю поступает на приемник, который иногда оформлен в виде платы, вставляемой в персональный компьютер. Цифровой сигнал с выхода приемника обрабатывается на компьютерах. Обработка включает в себя секторизацию, т.е. «вырезание» из всего спутникового изображения интересующего участка, например размером 512×512 пикселей, лежащего вблизи надир. Далее выполняются геометрическая коррекция изображения и топографическая привязка его к карте, а также коррекция атмосферных искажений. Секторизованное и скорректированное изображение готово для дальнейшей обработки, целью которой обычно является улучшение качества изображения, распознавание объектов на изображении, определение их координат и других геометрических характеристик.

1.4.3. Спутники для дистанционного зондирования

Спутник NOAA (США). Метеорологические и природоведческие спутники NOAA (рис. 1.23) имеют длину 4,18 м, диаметр 1,88 м, массу на орбите 1030 кг. Круговая орбита имеет высоту 870 км, один виток спутник совершает за 102 мин. Площадь солнечных батарей спутника $11,6 \text{ м}^2$, мощность батарей не менее 1,6 кВт, но со временем батареи деградируют из-за воздействия космических лучей и микрометеоров. Для нормальной работы спутника необходима мощность не менее 515 Вт.

В настоящее время на орбите функционируют несколько спутников. Сканер AVHRR спутника NOAA-14 с цилиндрическим сканированием имеет 8-дюймовую (20 см) оптическую систему Кассегрена, сканирование осуществляется путем вращения с частотой 6 об/с зеркала из бериллия. Угол сканирования $\pm 55^\circ$, полоса обзора около 3000 км. Из-за кривизны Земли зона радиовидимости спутника составляет ± 3400 км, поэтому за один проход спутника удается получить информацию с поверхности около 3000×7000 км.

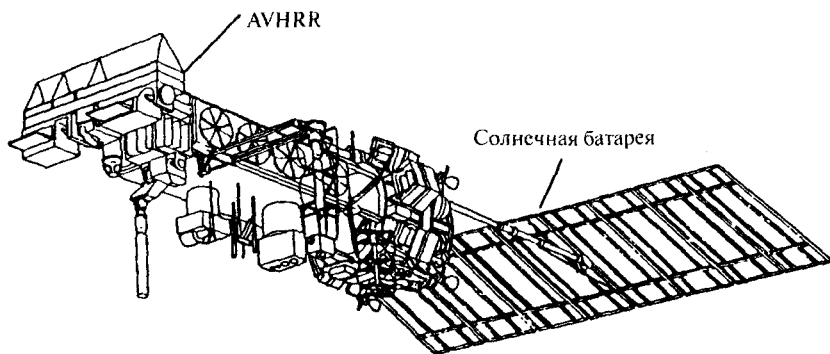


Рис. 1.23. Спутник NOAA

Спектральные каналы сканера выбраны так, что попадают в окна прозрачности атмосферы:

- 1 — 0,58–0,68 мкм (красный участок спектра);
- 2 — 0,725–1,0 мкм (ближний ИК);
- 3 — 3,55–3,93 мкм (участок ИК-диапазона, оптимальный для измерения излучения от лесных и других пожаров);
- 4 — 10,3–11,3 мкм (канал для измерения температуры поверхности суши, воды и облаков);
- 5 — 11,4–12,4 мкм (канал для измерения температуры поверхности суши, воды и облаков).

На спутнике NOAA-15 установлен дополнительный канал, работающий на волне длиной около 1,6 мкм для распознавания снега и льда.

В 1-м и 2-м каналах, спектральные характеристики которых приведены на рис. 1.24, в качестве детекторов излучения применяются кремниевые фотодиоды. В 4-м и 5-м каналах установлены охлаждаемые до

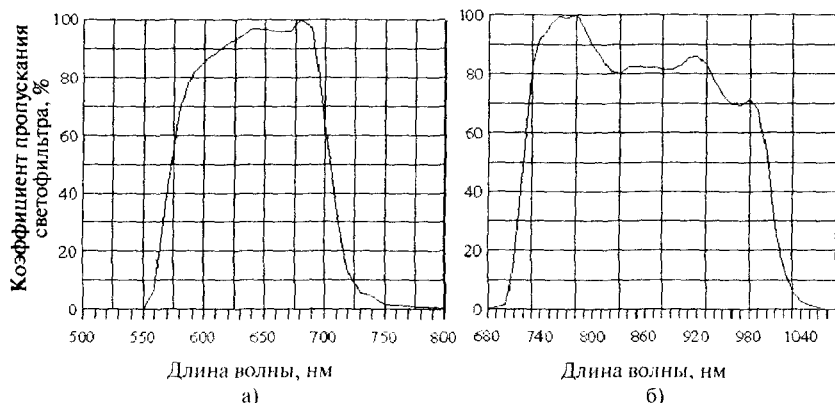


Рис. 1.24. Спектральные характеристики 1-го (а) и 2-го (б) каналов сканера AVHRR

105 К фоторезисторы на основе (HgCd) Те, в 3-м канале — охлаждаемый фоторезистор на основе InSb. На спутнике NOAA, как и на других спутниках, предусмотрена борговая калибровка датчиков.

Сканер AVHRR имеет мгновенное поле зрения во всех каналах $\Delta\varphi = 1,26 \cdot 10^{-3}$ рад, разрешение на местности в подспутниковой точке выбрано $\Delta L = 1,1$ км. Это связано с тем, что скорость спутника на орбите составляет 7,42 км/с, его проекция движется по поверхности Земли со скоростью 6,53 км/ч, сканер делает 6 сканов/с, за время одного скана проекция перемещается на $l = 6,53/6$ км = 1,09 км. Указанному полю зрения в подспутниковой точке соответствует пиксел $1,1 \times 1,1$ км. Сигналы каждого канала квантуются на 1024 уровня (10-битное квантование). Передатчик спутника имеет мощность 5,5 Вт, частота 1700 МГц. Скорость передачи цифровой информации со сканера AVHRR составляет 665,4 Кбит/с.

На спутнике установлена аппаратура HIRS для определения температуры в тропосфере на разных высотах (вертикальные профили атмосферы) в полосе обзора 2240 км. Для этого HIRS содержит автоматический сканирующий спектрофотометр ИК-диапазона, использующий свойство углекислого газа изменять положение и ширину линии поглощения на длинах волн порядка 14–15 мкм в зависимости от давления. Этот же прибор позволяет оценивать общее содержание озона ОСО в столбе атмосферы по поглощению теплового излучения от поверхности Земли и атмосферы на длине волны 9,59 мкм. И вертикальные профили, и ОСО вычисляются на приемном конце путем решения обратных задач.

Кроме указанной аппаратуры на спутник установлены: прибор SSU для исследования стратосферы; микроволновый прибор MSU для измерения температурных профилей стратосферы; аппаратура поиска и спасения по международной программе Коспас/SARSAT; система ARGOS для сбора метеорологической и океанографической информации с автоматических метеостанций, морских буев и воздушных шаров; некоторые другие приборы. ARGOS позволяет следить за миграцией крупных животных и птиц, если к их телу прикреплены специальные малогабаритные передатчики (рис 1.25).

Спутник «Ресурс-О1» (Россия). Высота орбиты 650 км, период обращения 97,4 мин, угол наклона орбиты 97°, 97. Сканер МСУ-СК с конической разверткой имеет скорость сканирования 12,5 дуг/с; разрешение 150×250 м; полоса обзора 600 км; спектральные каналы: 0,5–0,6 мкм (зеленый участок спектра), 0,6–0,7 мкм (красный участок), 0,7–0,8 мкм (красный и ближний ИК), 0,8–1,1 мкм (ближний ИК), 10,5–12,5 мкм (тепловой, в этом канале разрешение 500 м). Сигнал каждого канала квантуется на 256 уровней. Масса сканера 55 кг.

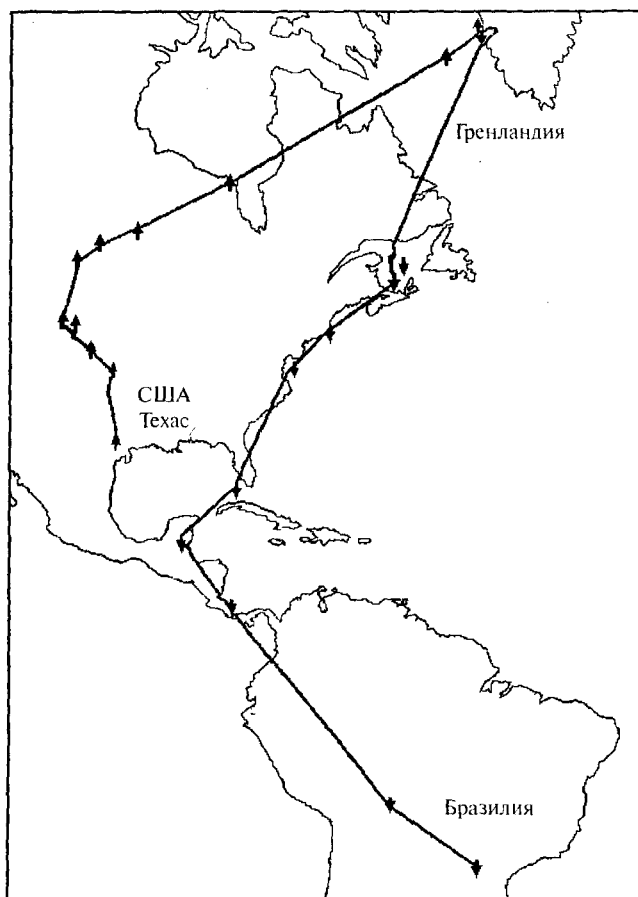


Рис. 1.25. Маршрут весеннего перелета (1995 г.) самца сокола-сапсана по данным ARGOS

На спутнике «Ресурс-О1» (рис. 1.26) установлены также два сканера МСУ-Э с линейной разверткой, содержащие по 3 линейки на ПЗС по 1000 пикселей (по одной на каждый из 3 спектральных каналов). Разрешение 35×45 м, скорость сканирования 200 строк/с; полоса обзора каждого сканера 45 км; если включены оба сканера, то полоса обзора составляет 80 км, так как полосы обзора перекрываются. Над одной и той же точкой поверхности спутник пролетает один раз в 14 дней. Чтобы повысить регулярность приема, предусмотрено отклонение оси сканера на $\pm 30^\circ$ от надира в направлении, перпендикулярном направлению движения спутника. Это позволяет смещать полосу обзора на ± 400 км. Спектральные каналы сканера: 0,5–0,59; 0,61–0,69; 0,7–0,89 мкм. Мас-

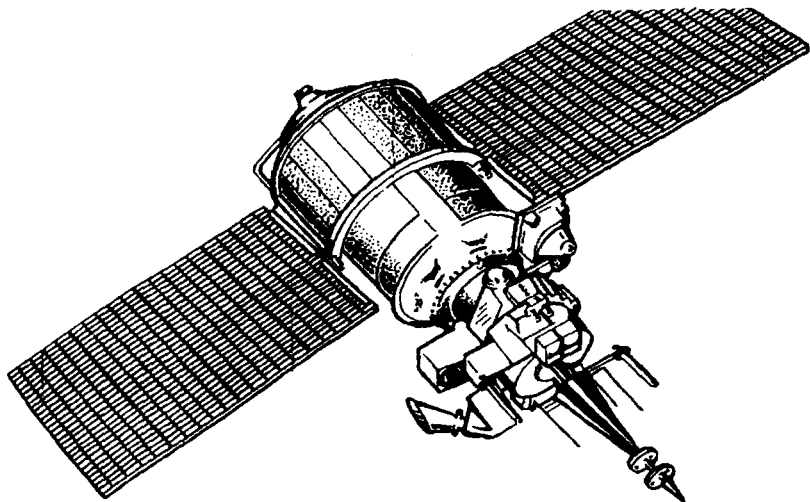


Рис. 1.26. Спутник «Ресурс-О1»

са прибора 23 кг. Результаты измерений передаются по радиоканалу на частоте около 8 ГГц со скоростью 7,68 Мбит/с, мощность бортового передатчика 10 Вт.

Спутник LANDSAT-5 (США). Высота орбиты 705 км, наклонение орбиты $98,2^\circ$, период обращения 98 мин. Над одной и той же точкой поверхности пролетает один раз в 16 дней приблизительно в 9 ч 45 мин местного времени. Установлены 2 сканера с цилиндрической разверткой: Multi-Spectral Scanner (MSS) и Thematic Mapper (TM). MSS имеет спектральные каналы 0,49–0,605 мкм (зеленый участок спектра), 0,603–0,7 мкм (красный), 0,701–0,813 мкм (красный — ближний ИК), 0,808–1,023 мкм (ближний ИК), разрешение $\Delta L = 80$ м, зона обзора 185×185 км. Сканирование осуществляется с помощью качающегося зеркала диаметром 30 см с частотой качания 13,62 Гц. Выходной сигнал квантуется на 64 уровня для каждого из каналов.

Thematic Mapper имеет разрешение $\Delta L = 30$ м во всех спектральных каналах, кроме шестого, где оно равно $\Delta L = 120$ м. 1–4-й каналы покрывают диапазон 0,45–0,9 мкм; 5-й — 1,55–1,75 мкм; 7-й — 2,08–2,35 мкм; 6-й канал тепловой (10,4–12,5 мкм). Формирование изображения осуществляется с помощью вращающегося зеркала диаметром 53 см с частотой 7 Гц. В 1–4-м каналах в качестве фотоприемников применяются кремниевые фотодиоды, в 5-м и 7-м каналах — фоторезисторы из InSb, охлаждаемые до 87 К, в 6-м канале использован фоторезистор из (HgCd) Te. TM имеет полосу обзора 185 км, выходной сиг-

нил каждого канала квантуется на 256 уровней, скорость формирования информационного потока 85 Мбит/с.

Если бы для каждого канала применялся один фотоприемник, то при указанных скоростях сканирования не удалось бы обеспечить указанного разрешения. Столь высокое разрешение сканеров достигнуто за счет применения линейки фотоприемников, ориентированной вдоль направления движения спутника, и последовательного считывания информации с элементов линейки.

Спутник SPOT-3 (Франция). Орбита почти круговая высотой 820 км с наклоном 98,7°, период обращения 101 мин. На спутнике установлены два сканера HVR с линейной разверткой, фотоприемниками служат 1728-элементные ПЗС-линейки, ориентированные перпендикулярно движению спутника, цветоделение осуществляется с помощью призм. Возможен панхроматический режим, перекрывающий зеленый, красный и ближний ИК-диапазоны (0,51–0,73 мкм). При этом используются 4 линейки, число пикселей в них достигает 6000, разрешение в надири $\Delta L = 10$ м.

В многоспектральном режиме реализуются три канала: 0,5–0,59, 0,61–0,68, 0,79–0,89 мкм, по две линейки на канал, $\Delta L = 20$ м. В панхроматическом режиме сигнал квантуется на 64 уровня, в многоспектральном — на 256 уровней для каждого канала. Полоса обзора равна 60 км в обоих случаях. Если включены оба сканера HVR, то полоса обзора равна 117 км с 3-километровым перекрытием. Периодичность наблюдения заданного района составляет 1 раз в 26 дней. Камера может отклоняться от направления в надири до $\pm 27^\circ$, поэтому полоса обзора может смещаться вправо и влево до 475 км от трассы спутника. При этом периодичность наблюдения заданного района может составлять 1–4 суток. Особенностью спутника SPOT является возможность получать стереоизображения земной поверхности путем съемки одного и того же участка на двух последовательных витках.

Спутник ERS (Европейское космическое агентство). Высота орбиты 798 × 782 км с наклоном 98,54° и периодом обращения 100,67 мин. В состав бортовой аппаратуры включена радиолокационная станция микроволнового зондирования AMI (Active Microwave Instrument), которая обеспечивает три режима работы.

Режим построения радиолокационных изображений подстилающей поверхности с использованием синтезированной апертуры антенны (AMI-SAR image mode) применяется при наблюдении береговой зоны, полярных льдов, при определении состояния поверхности моря, выявлении особенностей геологического строения земной поверхности, изучении растительного покрова. Сигналы, отраженные от поверхности Земли, могут приниматься двумя антеннами, расположенными од-

на над другой. По разности фаз их сигналов (интерферометрический метод измерения) можно определять высоту наземных объектов с точностью 10 м.

Аппаратура АМІ имеет следующие характеристики: мощность передатчика 1270 Вт, частота 5,3 ГГц, длительность импульса 37,1 мкс, ширина спектра излучаемых сигналов 15,5 МГц, вертикальная линейная поляризация излучаемых и принимаемых волн, пространственное разрешение 30 и 100 м, ширина полосы обзора 100 км при угле падения излучаемых электромагнитных волн на поверхность Земли 23° в центре полосы. На кодирование каждого отсчета отводится 16-битное слово.

Режим изучения морских волн с использованием синтезированной апертуры антенны (АМІ-SAR wave mode) обеспечивает определение направления и длины морских волн. Данный режим программно включается каждые 200–300 км, позволяя получать изображения размером 6×6 км, по которым можно оценивать характеристики морских волн. В этом режиме излучаемая мощность 540 Вт, частота 5,3 ГГц, поляризация излучаемых и принимаемых волн вертикальная, точность измерения длины волны $\pm 25\%$ в диапазоне 100–1000 м, точность определения направления волн $\pm 20^\circ$ в диапазоне 0–180°.

Режим трехлучевого скаттерометра (АМІ Scatterometer mode) предназначен для определения характеристик приповерхностных морских ветров. В этом режиме три передающие антенны формируют три луча, сканирующие в полосе шириной до 500 км, позволяя определять направление и скорость ветра. Элементы разрешения размером 50×50 км формируются с интервалом 25 км. Характеристики АМІ при этом следующие: излучаемая мощность 540 Вт, частота 5,3 ГГц, точность определения скорости ветра ± 2 м/с или 10% в диапазоне 2–24 м/с, точность определения направления ветра $\pm 20^\circ$ в диапазоне 0–360°, угол падения лучей изменяется в пределах 27–58°, обеспечивая полосу обзора шириной 500 км. Скорость передачи информации 500 кбит/с. Во всех трех режимах полоса обзора аппаратуры АМІ смещена относительно трассы ИСЗ вправо по ходу движения.

В состав измерительной аппаратуры входит также радиолокационный высотомер RA (Radar Altimeter) на частоту 13,8 ГГц для определения скорости ветра, измерения характерной высоты волн, топографии морской поверхности, ледяного покрова и поверхности суши, построения контуров ледяных массивов, а также выявления границ морских льдов. Высотомер может работать в режиме исследования океана (Ocean Mode), обеспечивая точность измерения скорости волн 2 м/с и точность измерения высоты волн 0,5 м в диапазоне 1–20 м в пределах пятна размером 1,6–2,0 км, точность определения высоты подъема поверхности моря 10 см.

В режиме определения характеристик ледяного покрова (Ice Mode) высотомер работает с более низким пространственным разрешением (около 7 км). Этот режим используется при топографии ледяных покровов, определении типа льда и выявлении границ ледяного покрова.

Для периодических уточнений параметров орбиты, а также калибровки радиолокационного высотомера на ERS устанавливаются лазерные уголкового отражатели.

Комплекс приборов ATSR (Along-Track Scanning Radiometer and Microwave Sounder) включает радиометр оптического диапазона и двухканальное микроволновое устройство вертикального зондирования. Радиометр предназначен для наблюдения поверхности моря и суши, измерения их температуры, температуры верхней облачности и обеспечивает прием излучения в спектральных каналах 0,65; 0,85; 1,27; 1,6; 3,7; 11 и 12 мкм с пространственным разрешением 1 км в надире.

Микроволновое устройство зондирования на частоты 23,8 и 36,5 ГГц обеспечивает поступление информации дистанционного зондирования о содержании аэрозолей в земной атмосфере, концентрации водяного пара и облачных капель, а также о состоянии растительного покрова Земли. Точность определения вертикального профиля концентрации водяного пара в пределах пятна размером 20–25 км составляет 10%.

Спектрометр GOME (Global Ozone Monitoring Experiment) используется для построения вертикальных профилей концентрации озона (O_3) и малых газовых компонентов (NO , NO_2 , BrO , H_2O) в тропосфере и стратосфере, измерения потоков солнечного излучения, отражаемого поверхностью Земли и рассеиваемого атмосферой. Прибор работает в ультрафиолетовом диапазоне в спектральных каналах 0,24–0,295; 0,29–0,405; 0,4–0,605 и 0,59–0,79 мкм. Каждый канал содержит решетку детекторов из 1024 фотодиодов, температура которых поддерживается в пределах 39–41 °С термоэлектрическими охладителями. Вертикальное разрешение при определении концентрации озона O_3 составляет 5 км. Полоса обзора изменяется от 120 до 960 км, а пространственное разрешение — от 40×40 до 40×320 км.

Аппаратура PRARE (Precise Range and Rate Equipment) обеспечивает определение параметров орбиты спутника путем одновременной передачи двух радиосигналов с разной частотой на сеть специальных наземных станций, измеренная разница времени прихода сигналов позволяет выполнить коррекцию относительной дисперсии, которая обусловлена влиянием ионосферы. Информация о дальности до спутника и его радиальной скорости передается обратно на борт спутника и накапливается в специальном бортовом запоминающем устройстве, а затем передается в пункт приема информации при пролете над ним. Точность определения наклонной дальности до спутника составляет 4–8 см.

Информация с аппаратуры АМІ, работающей в режиме построения радиолокационных изображений (АМІ-SAR image mode), передается со скоростью 105 Мбит/с на частоте 8,14 ГГц только в реальном масштабе времени. Информация со всей зондирующей аппаратуры космического аппарата ERS, за исключением радиолокационной, объединяется в цифровой поток со скоростью около 1,1 Мбит/с и передается на Землю на частоте 8,04 ГГц либо записывается на бортовой магнитофон для последующего воспроизведения в зоне радиовидимости станции приема информации со скоростью 15 Мбит/с и передачи на той же частоте.

Спутник «Океан-О» (Россия—Украина). Высота орбиты 667 км, период обращения 97,98 мин, угол наклона орбиты 98,03°. На спутнике установлены две некогерентные РЛС БО, антенны которых направлены соответственно влево и вправо от направления движения спутника. Полоса обзора каждой РЛС 455 км, угол падения излучаемых электромагнитных волн на подстилающую поверхность 20,48°. Пространственное разрешение в среднем 1,3 км поперек направления движения спутника и 2,5 км вдоль него.

Многоканальный сканирующий микроволновый радиометр «Дельта-2Д» имеет 4 канала со средней длиной волны 0,8 мкм (1-й канал), 1,35 см (2-й канал), 2,25 см (3-й канал) и 4,3 см (4-й канал). Полоса обзора составляет 1126 км, размеры элементов разрешения на подстилающей поверхности не более 17×22 км в 1-м канале, 28×37 км во 2-м, 49×65 км в 3-м и 91×120 км в 4-м.

Два микроволновых несканирующих радиометра на длины волн 2,25 и 6 см направлены в надиr, имеют полосу обзора около 130 км.

На спутнике «Океан-О» установлен ряд приборов, работающих в оптическом диапазоне. Сканер МСУ-СК с конической разверткой аналогичен сканеру МСУ-СК спутника «Ресурс-О». Многоканальное сканирующее устройство высокого разрешения МСУ-В с полосой обзора 195 км имеет 8 спектральных каналов:

- 1-й — 0,48–0,52 мкм;
- 2-й — 0,54–0,61 мкм;
- 3-й — 0,63–0,73 мкм;
- 4-й — 0,78–0,92 мкм;
- 5-й — 0,915–0,997 мкм;
- 6-й — 1,47–1,62 мкм;
- 7-й — 2,060–2,385 мкм;
- 8-й — 10,6–12,0 мкм.

Пространственное разрешение в 1–5-м каналах 50 м, в 6-м — 100 м, в 7-м — 300 м, в 8-м — 250 м.

Многоканальное сканирующее устройство малого разрешения МСУ-М имеет полосу обзора 2000 км, пространственное разрешение $1,5 \times 1,7$ км в центре строки и $1,8 \times 1,7$ км на краю строки.

Поляризационный спектрорадиометр «Трассер-О» в диапазоне длин волн 411–809 нм имеет пространственное разрешение 45 км.

Для сбора информации с морских буев предназначена аппаратура сбора и передачи информации «Кондор-2».

Кроме России, США, Франции, Украины, Европейского космического агентства, спутники для дистанционных исследований Земли запущены Японией, Индией, Китаем, Бразилией и некоторыми другими странами.

1.4.4. Искажения спутниковых сканерных изображений

Рассмотрим схему формирования спутниковых изображений с учетом кривизны Земли. Пусть спутник огибает поверхность Земли по круговой орбите на высоте H . Это равносильно движению спутника по прямой, но над поверхностью цилиндра. Кажущееся изображение, которое строит сканер спутника с цилиндрической или линейной разверткой, — это проекция на поверхность цилиндра F (рис. 1.27).

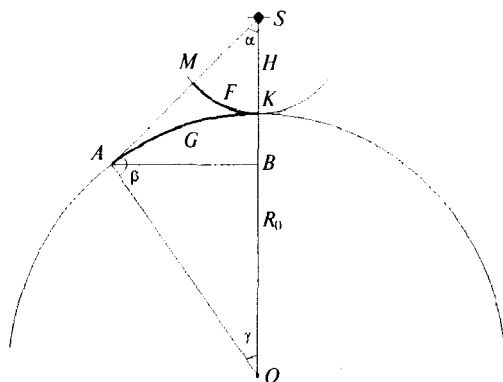


Рис. 1.27. Диаграмма формирования спутниковых сканерных изображений

В действительности же сканер отображает земную поверхность, т.е. внешнюю поверхность цилиндра G . Будем считать, что Земля — шар радиусом $R_0 = 6370$ км; собственное вращение Земли за время построения сканерного изображения учитывать не будем.

Пусть $L = SA$ — расстояние от спутника S до некоторой точки земной поверхности при сканировании, ΔL — разрешение в надире, $\Delta \varphi$ — мгновенное поле зрения, α — угол сканирования (угол визирования), под углом γ дуга AK видна из центра Земли. Используя теорему синусов и учитывая, что для всех значений угла α выполняется соотношение $(H + R_0) \sin \alpha < H$, получаем

$$SO / \sin \beta = AO / \sin \alpha; \sin \beta = (1 + H/R_0) \sin \alpha; \gamma = \pi - \beta - \alpha;$$

$$\gamma = \pi - \arcsin[(1 + H/R_0)\sin \alpha] - \alpha; L = R \sin \gamma / \sin \alpha.$$

Длина дуги AK , отвечающей развертке внешней поверхности цилиндра G , равна $y = AK = R_0\gamma$. Длина дуги MK , соответствующей кажущемуся изображению, $x = MK = H\alpha$, откуда $\alpha = x/H$. Таким образом,

$$y = R\{\pi - \arcsin [(1 + H/R_0) \sin (x/H)] - x/H\}. \quad (1.15)$$

Используя (1.15), можно пересчитать координату x вдоль строки кажущегося изображения в координату вдоль строки реального изображения.

Знание расстояния L позволяет уточнить пространственное разрешение. Минимальная дальность $L_{\min} = H$, при этом разрешение $\Delta L = \Delta\phi H$; при максимальном угле сканирования α_{\max} дальность равна L_{\max} , разрешение $\Delta L_1 = \Delta\phi L_{\max}$ вдоль направления движения спутника. Но поверхность видна со спутника под углом α_{\max} , поэтому поперек направления движения $\Delta L_2 = \Delta L_1 / \cos \alpha_{\max}$. Таким образом, на максимальной дальности поперек направления движения пиксеты оказываются сжатыми в $\Delta L_2 / \Delta L$ раза, вдоль — в $\Delta L_1 / \Delta L$ раза. Кажущаяся полоса обзора $F_{\max} = H\alpha_{\max}$, реальная полоса $G_{\max} = 2R\beta_{\max}$.

Для спутника NOAA $H = 870$ км, $\alpha_{\max} = 55^\circ$, $\Delta\phi = 1,26 \cdot 10^{-3}$ рад, $\Delta L = 1,1$ км. Максимальная дальность $L_{\max} = 1827$ км, полоса обзора $G \approx 3000$ км, разрешение $\Delta L_1 = 2,31$ км и $\Delta L_2 = 4,02$ км. На максимальной дальности поперек направления движения пиксеты оказываются сжатыми в 3,65 раза, вдоль него — в 2,1 раза. Две реки, текущие параллельно направлению движения спутника и находящиеся на расстоянии 1,1 км друг от друга, в надире были бы различимы, но на краю скана они сливаются. Если не проводить коррекцию, то полоса обзора кажется равной $F = 2H\alpha_{\max} = 1670$ км.

На рис 1.28 показаны кажущееся и действительное изображения двух сканов. За время одного скана спутник NOAA перемещается на

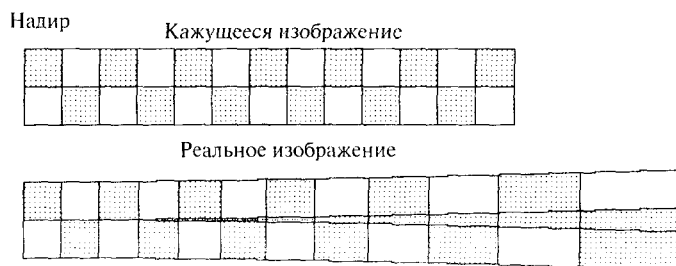


Рис. 1.28. Кажущееся и реальное изображения двух строк сканирования (спутник NOAA)

1,1 км при разрешении в надире 1,1 км. На некотором расстоянии от надира строки изображения начинают заметно перекрываться.

Определим область спутникового изображения с малыми искажениями как участок, где строки перекрываются не более чем на 10%, т.е. примерно на 0,1 км. При этом $\Delta L_1 = 1,3$ км. Этот участок соответствует кажущейся полосе обзора в ± 475 км и действительной полосе обзора в ± 540 км от надира. Разрешение вдоль скана на границе участка составляет $\Delta L_2 = 1,54$ км, вдоль скана пиксели вытянуты в 1,4 раза.

Для спутников с малой полосой обзора (спутник «Ресурс-О1-3», сканер МСУ-Э; спутник SPOT, сканер HVR) при наблюдении в надир геометрические искажения изображений за счет кривизны Земли незначительны. Искажения возникают при отклонении оси сканирования от надира, которое может достигать около 30° (*перспективные искажения*). Это иллюстрирует рис. 1.29, построенный на основании (1.15), где в центре показана полоса шириной 45 км — область кажущегося изображения при отклонении оси сканера МСУ-Э на 30° . В действительности изображение несколько вытянуто вдоль направления сканирования (60,5 км). При этом разрешение вдоль направления движения спутника составляет $\Delta L_1 = 41$ м на максимальной дальности и $\Delta L_1 = 39$ м на минимальной, в то время как в надире $\Delta L_1 = 35$ м. В надире для МСУ-Э разрешение вдоль линии сканирования $\Delta L_2 = 45$ м, при отклонении оси сканирования на 30° $\Delta L_2 = 60$ м.

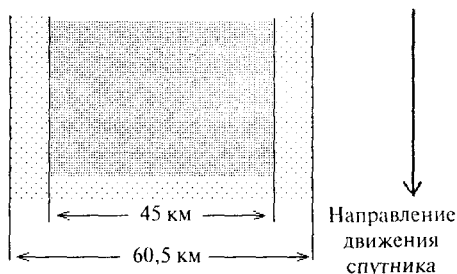


Рис. 1.29. Искажение изображения со спутника «Ресурс-О1-3» при отклонении оси сканирования на 30° от надира

Сканер МСУ-СК спутника «Ресурс-О1-3» имеет спирально-коническую развертку, благодаря чему сканирующий луч описывает по поверхности Земли дугу. Так как расстояние от спутника до каждой точки дуги одно и то же, то искажения сканерного изображения за счет кривизны Земли не возникают. Строки на изображении также следует строить в виде дуг. Однако такое изображение передается со спутника в том же формате, что и изображения от сканеров с цилиндрической и линейной разверткой, у которых строки — прямые линии. Переход от линейного формата к формату, учитывающему особенности сканера со спирально-конической разверткой, при известных параметрах сканера несложен.

Рассмотренные причины геометрических искажений спутниковых изображений не единственно возможные. Солнечно-синхронные орбиты природоведческих спутников не пересекают ось вращения Земли, а наклонены относительно нее. Поэтому если спутник движется с севера на юг (нисходящий виток орбиты), то вверху изображения будет не север, как на карте, а, например, север-северо-восток. К тому же во время сеанса приема спутниковой информации Земля поворачивается на некоторый угол (на $0,25^\circ$ за 1 мин). Кроме «свежих» сканерных изображений, приходится обрабатывать архивные изображения, для которых неизвестны и орбитальные данные, и угол отклонения оси сканирования от надира. В этих случаях необходим другой метод геометрической коррекции (см. п. 3.1.2).

КОНТРОЛЬНЫЕ ЗАДАНИЯ

1. На какую длину волны приходится максимум теплового излучения от лесного пожара, если горящий лес имеет температуру 600°C ?
2. Используя рис. 1.5, оцените вегетационный индекс для кукурузы и сои.
3. Найдите на рис. 1.6 линии поглощения озона и углекислого газа в ИК-диапазоне.
4. Пусть оптическая толщина атмосферы $\tau = 0,2$. Во сколько раз отличаются интенсивности принимаемого аппаратурой спутника излучения при наблюдении в надири и под углом $\delta = 30^\circ$?
5. Каково мгновенное поле зрения сканеров спутников NOAA (высота орбиты $H = 870$ км) и «Ресурс-О1-3» ($H = 650$ км)?
6. Карте какого масштаба соответствует панхроматическое изображение со спутника SPOT с разрешением 10 м?
7. Каковы период обращения и орбитальная скорость японского спутника исследования природных ресурсов Jers-1, если высота орбиты составляет 568 км?
8. С какой скоростью движется по поверхности Земли проекция спутника, если высота орбиты 900 км?
9. Солнечная постоянная на орбите $1,4 \text{ кВт/м}^2$, КПД солнечных батарей 0,15. Какова мощность новых солнечных батарей спутника NOAA?
10. Подсчитайте скорость спутника NOAA на орбите.
11. Сколько надо делать сканов в секунду, чтобы обеспечить разрешение в 50 м с полосой обзора 2400 км? Какова должна быть скорость передачи информации в этом случае?
12. Какова скорость спутника «Ресурс» на орбите и скорость движения проекции спутника на земную поверхность? С какой частотой следует качать зеркало сканера, чтобы получить разрешение в 45 м при использовании сканера с качанием зеркала вместо ПЗС-линейки?
13. Какого размера зеркальный объектив следует установить на геостационарном спутнике, чтобы получить разрешение в 10 м в зеленом участке спектра?

14. Подсчитайте скорость передачи информации со сканера МСУ-Э спутника «Ресурс-О1» при разрешении 35×45 м.

15. Подсчитайте число элементов в фотоприемной линейке каждого канала сканеров MSS и ТМ спутника LANDSAT.

16. Орбитальные элементы в формате NORAD спутника OCEANSAT (Индия) имеют вид

1 25758U 99029C 99351.20895206.00001054 00000-0 27840-3 0 1586
2 25758 98.2916 85.3537 0001002 80.8953 279.2357 14.50869716 29712

К какому моменту времени относятся орбитальные элементы, каковы угол наклона орбиты, долгота восходящего узла, эксцентриситет орбиты, аргумент перигея, средняя аномалия, среднее движение в оборотах за сутки? Вычислите период обращения спутника в минутах и высоту орбиты в километрах.

17. Какой размер должна иметь антенна для приема потока информации в 100 Мбит/с со спутника, находящегося на расстоянии 2000 км, если шумовая температура МШУ составляет 100 К, а мощность передатчика 10 Вт?

18. Каков максимальный угол сканирования прибора МСУ-СК спутника «Ресурс-О1-3», если полоса обзора составляет 600 км? Чему равно разрешение в надире и на границе полосы обзора?

19. Чему равно разрешение, даваемое прибором MSS спутника LANDSAT на границе полосы обзора?

20. К каким дополнительным искажениям изображения приводит вращение Земли в течение 15-минутного сеанса приема информации со спутника NOAA на средних широтах?

ЛИТЕРАТУРА

Кондратьев К.Я., Тимофеев Ю.М. Метеорологическое зондирование атмосферы из космоса. Л.: Гидрометеиздат, 1978. 279 с.

Зуев В.Е., Креков Г.М. Оптические модели атмосферы. Л.: Гидрометеиздат, 1986. 256 с.

Хргиан А.Х. Физика атмосферы. М.: Изд-во МГУ, 1988. 327 с.

Гарбук С.В., Гершензон В.Е. Космические системы дистанционного зондирования Земли. М.: Сканэкс, 1997. 296 с.

Киенко Ю.П. Введение в космическое природоведение и картографирование. М.: Картгеоцентр— Геодезиздат, 1994. 214 с.

Дистанционное зондирование: количественный подход: Пер. с англ. / Под ред. А.С. Алексеева. М.: Недра, 1983. 415 с.

Брюханов А.В., Господинов Г.В., Книжников Ю.Ф. Аэрокосмические методы в географических исследованиях. М.: Изд-во МГУ, 1982. 231 с.

Кац Я.Г., Тевелев А.В., Полетаев А.И. Основы космической геологии. М.: Недра, 1988. 236 с.

Гонин Г.Б. Космическая фотосъемка для изучения природных ресурсов. Л.: Недра, 1989. 320 с.

Криксунов Л.З., Падалко Г.А. Тепловизоры. Киев: Техніка, 1987. 166 с.

Радиолокационные методы исследования Земли / Под ред. Ю.А. Мельни-
ка. М.: Сов. радио, 1980. 262 с.

Комаров С.А., Миронов В.Л. Микроволновое зондирование почв. Новоси-
бирск: Наука, 2000. 280 с.

Астрономический календарь. Постоянная часть. М.: Наука, 1981. 704 с.

2. ПРЕДСТАВЛЕНИЕ ИЗОБРАЖЕНИЙ В ЭВМ

2.1. ВИДЫ ИЗОБРАЖЕНИЙ

На ранних стадиях развития вычислительной техники компьютеры использовались для обработки полученных в ходе эксперимента сигналов, для проведения различных вычислений и для математического моделирования. В последние десятилетия все шире применяется цифровая обработка изображений. Это очень важно, поскольку для человека зрение — самый естественный механизм восприятия информации о внешнем мире. Компьютеры позволяют преобразовывать изображения, устранять их искажения и даже «делать невидимое видимым». Одно из главных направлений цифровой обработки изображений — автоматизированное распознавание образов. В настоящее время компьютер стал стандартным средством обработки изображений при дистанционных исследованиях, в робототехнике, в издательском деле, материаловедении, в медицине, криминалистике. Компьютер успешно справляется с обработкой видеoinформации в реальном времени, позволяет создавать новые изображения (машинная графика). Различными фирмами выпускается обширный спектр технических и программных средств обработки изображений, ввода и вывода их из ЭВМ.

Будем называть изображением функцию двух переменных $f(x, y)$, определенную в некоторой области S плоскости Ox, y и имеющую известное множество своих значений. Например, обычную черно-белую фотографию можно представить как $f(x, y) \geq 0$, $0 \leq x \leq a$, $0 \leq y \leq b$, где $f(x, y)$ — яркость (иногда называемая оптической плотностью или степенью белизны) изображения в точке (x, y) ; a — ширина; b — высота кадра. С учетом особенностей функции f обычно выделяют следующие классы изображений:

1. Полутоновые (серые) изображения (рис. 2.1 а). Примерами таких изображений могут служить обычные черно-белые фотографии. Множество значений функции в S может быть дискретным ($f \in \{f_0, \dots, f_l\}$, $l > 1$) либо непрерывным ($0 \leq f \leq f_{\max}$). В один класс с полутоновыми обычно относят цветные фото- и телевизионные изображения.

Экспериментально установлено, что практически любой цвет, видимый человеком, но не всякий, может быть представлен как сумма трех линейно независимых так называемых основных цветов, взятых с весами m_1, m_2, m_3 :

$$f = m_1 \cdot E_1 + m_2 \cdot E_2 + m_3 \cdot E_3.$$

В соответствии с принятой системой RGB основными являются цвета, соответствующие монохроматическим излучениям с длиной волны: $\lambda = 0,7$ мкм (красный — R); $\lambda_2 = 0,5461$ мкм (зеленый — G); $\lambda_3 = 0,4358$ мкм (голубой — B). Здесь множество значений функции f представляет собой 3-мерное пространство интенсивностей (m_1, m_2, m_3) основных цветов. Если $m_1 = m_2 = m_3 = m$, то в зависимости от m получаются различные оттенки серого цвета — от черного до белого. Если $m_1 \neq m_2 \neq m_3$, то изображение окрашено в различные цвета.

При космической и аэро съемке применяют многоспектральную фото-, теле- и сканерную аппаратуру, спектральные каналы которой могут лежать как в видимом, так и в невидимых УФ- и ИК-участках спектра; число таких каналов может достигать $n = 5-10$ и более. Здесь множество значений функции f представляет собой n -мерное пространство интенсивностей (m_1, m_2, \dots, m_n). При визуализации таких изображений на мониторах ЭВМ каждому вектору этого пространства можно присвоить какой-либо видимый человеческим глазом цвет, не обязательно основной. В результате получим аэро- или космическое изображение *в условных цветах*.

Человеческий глаз не отличает оттенки серого изображения, яркость которых различается менее чем на 1–2%, но хорошо видит цвет. С помощью компьютера разные уровни яркости полутонового изображения можно сопоставить с некоторой цветной шкалой и окрасить изображение в так называемые *ложные цвета*. Правильно подобранная шкала позволяет оператору визуально выделять объекты, мало отличающиеся по яркости.

2. Бинарные (двухуровневые) изображения (рис. 2.1, б). Здесь множество значений функции $f \in (0, 1)$. Обычно 0 — это черный цвет, а 1 — белый. Примером бинарного изображения являются карты водоемов, карты плоских деталей на конвейере и т.д.

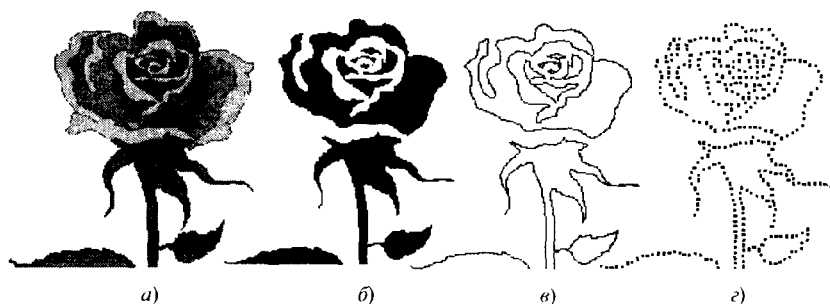


Рис. 2.1. Виды изображений: полутоновое (а), бинарное (б), линейное (в), точечное (г)

3. Линейные изображения (рис. 2.1, в). Типичный пример — карта изолиний, изображение береговой линии, различные контуры. Изображение представляет собой одну кривую или семейство их.

4. Точечные изображения (рис. 2.1, г). Типичный пример — кадр участка звездного неба. Изображение представляет собой k точек с координатами (x_i, y_i) и яркостью f_i , $i = 1, \dots, k$.

Все многообразие задач обработки изображений можно представить, как преобразования между классами изображений, а также внутри класса. Так, понятие «улучшение качества» обычно относят к преобразованиям полутоновых (или цветных) изображений; сегментация состоит в преобразовании полутонового изображения в бинарное; линейные изображения получают, оконтуривая бинарные изображения; кривые и линии преобразуют в точечные изображения, отыскивая критические точки контуров (в случае многоугольников этими точками являются вершины). Критические точки обычно используются в задачах машинного зрения как признаки образов при их распознавании. При обработке изображений, получаемых при дистанционном зондировании Земли из космоса, критические или опорные точки (обычно это элементы гидросети) используются для совмещения изображения с картой, при коррекции геометрических искажений изображения.

Можно считать, что преобразование изображений из класса с меньшим номером в класс с большим номером относится к задачам сжатия изображений и распознавания образов. После операции распознавания, когда объект идентифицирован, часто бывает необходимо провести измерение геометрических и других параметров объектов на изображении. Например, по космическому снимку можно оценить площадь гари после лесного пожара и координаты его центра тяжести. Для определения площади необходимо получить бинарное изображение, где объект (пожарище) белого цвета, а фон — черного (или наоборот). После нахождения указанных координат бинарное изображение пожарища превращается в точечное, где точка — центр тяжести.

Преобразование изображений класса 4 в изображения класса 3 связано с интерполированием и аппроксимацией, в результате чего через заданные точки проводится гладкая кривая. Преобразование изображений класса 3 в изображения класса 2 называется заполнением контура. Иногда преобразования изображений класса 2 в изображения класса 1 возможны путем сглаживания бинарного изображения или подмешивания низкочастотного шума. Преобразования изображений из класса с большим номером в класс с меньшим номером обычно относят к задачам машинной графики. Важным современным направлением обработки изображений является восстановление трехмерного изображения по его двумерным проекциям — компьютерная томография.

весьма значимая в медицине и в технологическом контроле в промышленности.

Регистрация, компьютерная обработка и воспроизведение изображений предполагают их организацию в виде двумерного массива f_{ij} , $i = 1, \dots, n$; $j = 1, \dots, m$, где f_{ij} — яркость пиксела (элемента изображения) в i -й строке и j -м столбце. Такая организация используется в датчиках изображений — сканерах авиационного и космического базирования, передающих телевизионных трубках, твердотельных передающих приборах, ручных и настольных сканерах для ввода изображений с фотографий и рисунков. Рассмотренный способ представления изображений называется *растровой графикой*. Растровая графика используется также в телевизионных приемных трубках, в мониторах ЭВМ.



Рис. 2.2. Фрагмент растрового изображения

На рис. 2.2 несколько строк растрового изображения показаны в увеличенном виде, видны отдельные пикселы.

2.2. ДАТЧИКИ ИЗОБРАЖЕНИЙ

2.2.1. Передающая телевизионная трубка — видикон

Действие видикона основано на явлении внутреннего фотоэффекта (фотопроводимости). На внутренней поверхности входного окна видикона нанесен тонкий слой полупроводникового фотопроводника — мишень. Каждый элемент мишени можно рассматривать как параллельное соединение конденсатора и светочувствительного резистора R . Чем выше освещенность соответствующего элемента, тем меньше R . Электронный луч (пучок), излучаемый катодом, обегает мишень и заряжает каждый конденсатор до потенциала катода. После ухода луча конденсатор разряжается через резистор тем быстрее, чем меньше сопротивление R . Подзарядка конденсаторов, т.е. компенсация утечки их зарядов электронным лучом, при последующем сканировании сопровождается протеканием зарядного тока в цепи катода; ток, таким образом, пропорционален освещенности элементов мишени. Видикон чувствителен в видимом диапазоне спектра (0,4–0,8 мкм). Для регистрации цветного RGB-изображения используются три видикона, перед каждым из которых установлен соответствующий светофильтр.

Трубки типа видикон с мишенью из пироэлектрика (пирикены) применяются для наблюдения изображений в тепловом диапазоне (8–14 мкм) и позволяют оценивать температуру различных участков объекта на изображении. Пироэлектричество — явление возникновения электрических зарядов на поверхности некоторых кристаллов при их нагревании или охлаждении. Тепловое излучение наблюдаемого объекта фокусируется на мишень из пироэлектрика (например, из триглицинсульфата) и формирует на ее поверхности температурный рельеф, который в свою очередь вызывает на этой поверхности дополнительные электрические заряды, суммируемые с зарядом, наводимым электронным пучком. Далее движущийся электронный луч проходит мишень — строка за строкой, причем сила тока пропорциональна дополнительному заряду. Таким образом, ток на выходе пирикона отражает карту распределения температуры объекта. Сигнал с выхода пирикона удобнее всего вводить в ЭВМ и далее обрабатывать, для визуализации раскрашивая в ложные цвета.

Недостатками вакуумных электронно-лучевых передающих трубок являются использование высоких напряжений (до 1000 В) и сравнительно небольшой срок службы при значительных габаритах.

2.2.2. Твердотельные датчики

В настоящее время наиболее распространены твердотельные фоточувствительные линейки и матрицы, выполненные на приборах с зарядовой связью (ПЗС), регистрирующих излучение в видимом и ближнем инфракрасном диапазонах (0,4–1,1 мкм). В линейном приборе на ПЗС (линейке ПЗС) фоточувствительные элементы (пикселы) расположены в один ряд, их число доходит до 1000 и более при поперечном размере каждого элемента 10–20 мкм.

ПЗС-матрица — это совокупность линеек, число которых может достигать нескольких сотен. Она представляет собой кремниевую пластину размером 1–2,5 см, на которой расположены несколько сот тысяч фоточувствительных элементов, регистры переноса зарядов и схема управления. В качестве фоточувствительных элементов используются дискретные конденсаторы со структурой металл–окисел–полупроводник (МОП). На поверхность ПЗС-линейки или матрицы с помощью объектива проецируется изображение наблюдаемого объекта (рис. 2.3). Поглощенные фотоны вызывают генерацию электронно-дырочных пар, неосновные носители накапливаются в МОП-конденсаторах. Накопленный заряд пропорционален числу попавших фотонов, т.е. освещенности конкретного МОП-конденсатора. Изменяя смещение на электродах МОП-конденсаторов, заряды перемещают поочередно к вы-

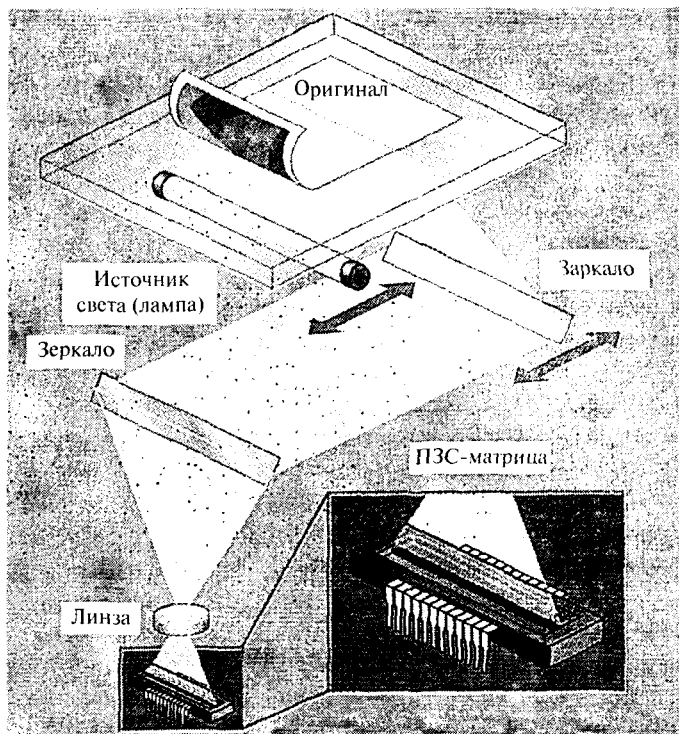


Рис. 2.3. Планшетный сканер

ходу прибора на ПЗС, осуществляя электронное сканирование. На выходе заряд детектируется как переменный по времени видеосигнал. Для регистрации цветного RGB-изображения на фоточувствительные элементы наносят пленочные светофильтры в виде цветных полос или цветной мозаики, возможно применение цветоделительных призм.

Фоточувствительные матрицы используются в твердотельных теле- и видеокамерах, а фоточувствительные линейки ПЗС — в устройствах считывания изображений с плоских оригиналов — ручных и настольных сканерах, в которых линейка перемещается относительно оригинала, в том числе некоторых сканерах космического базирования. В последнее время появились фоточувствительные матрицы со структурой, сходной с ПЗС, которые работают в среднем и тепловом участках ИК-диапазона. Кроме того, для ввода изображений применяют точечные фотоприемники.

Сканеры космического базирования с механическим сканированием (вращающееся или качающееся зеркало), работающие в видимом

и ближнем ИК-диапазонах (0,3–1,2 мкм), построены на кремниевых фотодиодах (ФД), в среднем и дальнем ИК — на фоторезисторах (ФР). Фотодиод закрыт приложенным к нему напряжением, и ток основных носителей через него практически отсутствует; под действием кванта света в ФД возникают неосновные носители, для которых это напряжение, наоборот, является ускоряющим. Ток неосновных носителей пропорционален освещенности. Кремниевые фотодиоды, из которых можно сформировать фотодиодные линейки и матрицы, обладают малой инерционностью.

Фоторезистор представляет собой полупроводник, заключенный между двумя электродами. Под действием квантов света увеличивается число основных носителей — свободных электронов или дырок, возрастает проводимость. ФР обладают высокой чувствительностью, однако существенно более инерционны, чем ФД и ПЗС. В диапазоне 3–5 мкм используются ФР на InSb, в диапазоне 8–12 мкм — ФР на (CdHg)Te. Для уменьшения темнового тока и его флуктуации ФР охлаждаются, например термоэлектрическими микрохолодильниками, работа которых основана на эффекте Пельтье.

2.3. ВВОД ИЗОБРАЖЕНИЙ В ЭВМ

Рассмотренные датчики позволяют получать полутоновые и цветные изображения в аналоговой форме. Для ввода в ЭВМ их необходимо преобразовать в некоторую числовую матрицу. Процесс преобразования называется дискретизацией и состоит из выборки и квантования. Первая заключается в выборе на поле наблюдения некоторого множества точек, в каждой из которых измеряется яркость $f(x, y)$. При использовании ПЭВМ для обработки изображений часто применяется формат 512×512 , так что упомянутое множество точек состоит из $512 \cdot 512 = 262\,144$ пикселей. Выборка производится с некоторым шагом, который и характеризует пространственное разрешение. При использовании сканеров космического базирования разрешение обычно определяется скоростью спутника и частотой сканирования.

В настоящее время основным устройством ввода изображений в ЭВМ с плоских оригиналов является планшетный сканер (см. рис. 2.3). В практике дистанционных исследований он применяется для ввода топографических карт. Типичный планшетный сканер содержит источник света и систему зеркал (одно из них вместе с лампой перемещается посредством шагового двигателя вдоль сканируемого изображения). Световой поток попадает через линзовый объектив в ПЗС-матрицу, которая обычно состоит из трех ПЗС-линеек с пленочными или матричными

светофильтрами. Оптическое разрешение по строке определяется размером ПЗС-матрицы и составляет до 2400 точек на дюйм, в перпендикулярном направлении — числом шагов развертки (до 1200 точек на дюйм). Предусмотрено программное обеспечение для повышения разрешения путем интерполяции.

Для квантования и преобразования $f(x, y)$ в числовой код в устройствах ввода используется аналого-цифровой преобразователь (АЦП). Поскольку человеческий глаз обычно не в состоянии различать уровни освещенности, различающиеся менее чем на 1–2%, для качественного представления полутонового изображения вполне достаточно иметь $2^8 = 256$ уровней квантования, т.е. по 1 байту на пиксел. Это значение характерно для *тонового разрешения*. При дискретизации цветного изображения обычно затрачивают по 1 байту на каждый из трех основных цветов, т.е. 3 байта на пиксел (*24-битное кодирование, 24-битная глубина цвета*). Черно-белые сканеры регистрируют не менее 256 градаций серого, цветные — не менее 256 градаций каждого из трех основных цветов, т.е. не менее 16,7 млн оттенков. В некоторых системах, например в сканерах космического базирования, число уровней квантования выбирается исходя из конкретной задачи.

В барабанных сканирующих устройствах изображение (обычно это фотонегатив) закрепляется на вращающемся барабане, параллельно оси которого перемещается сфокусированный световой пучок, освещающий небольшой (например, диаметром 10 мкм) участок изображения. Прошедший через негатив свет принимается точечным фотоприемником и преобразуется в электрический сигнал, который далее подается на АЦП. Такие сканирующие устройства обеспечивают самое высокое качество ввода изображений в ЭВМ.

К числу устройств, специально предназначенных для ввода изображений в ПЭВМ, относятся цифровые фото- и видеокамеры. Типичная цифровая фотокамера содержит ПЗС-матрицу с разрешением 1280×960 пикселей, АЦП и процессор, изображение в цифровой форме записывается на жесткий или флоппи-диск, для хранения изображений может также использоваться полупроводниковая память. В ПЭВМ изображение вводится либо по кабелю, либо через бесконтактное устройство ввода на ИК, либо дискетой. По стандарту Portable PC Camera ПЗС-матрица имеет разрешение 640×480 пикселей. Цифровые видеокамеры формата DV (Digital Video) имеют разрешение 767×575 пикселей, изображения в цифровой форме записываются на специальные кассеты с магнитной пленкой.

По качеству изображения твердые копии небольшого формата, полученные с помощью цифровой камеры, не уступают обычной фотографии размером 9×13 см, однако 35-мм фотонегативы обычно допуска-

ют 10-кратное увеличение без существенной потери качества; на увеличенном в 3—4 раза изображении с цифровой камеры просматривается пиксельная структура.

Ввод изображений в ПЭВМ с телевизора, аналогового видеомagnetofона, видеокамер с аналоговой записью формата VHS, Super-VHS и др. осуществляют с помощью видеоплаты с АЦП. Максимальное разрешение, которое обеспечивает обычное телевидение, равно 767×575 пикселей. Видеомagnetитофоны и камеры VHS позволяют получить разрешение не более чем 500×380 пикселей, Super-VHS — до 640×480 пикселей.

Рассмотренные устройства не являются единственными известными устройствами регистрации и ввода изображений в ЭВМ. В медицине, например, используются ультразвуковые приборы, в которых излучателем и приемником являются пластины из пьезоэлектрика. Источником излучения могут быть рентгеновская трубка и радиоизотопный источник, а приемником — фотопленка, телевизионная камера, сканер с точечным детектором квантов. Строение тканей организма по распределению в них протонов оценивают сканером, использующим ядерный магнитный резонанс. Нередко с помощью компьютера строят не двумерные, а трехмерные изображения (томография).

2.4. УСТРОЙСТВА ВИЗУАЛИЗАЦИИ ИЗОБРАЖЕНИЙ

Для визуализации изображений используют цветные мониторы на электронно-лучевых трубках и жидкокристаллических дисплеях. Современные электронно-лучевые мониторы имеют трубки с диагональю 15, 17, 20 дюймов и более, разрешение 15-дюймового монитора составляет 1024×768 пикселей при размере зерна экрана 0,24—0,27 мм, 20-дюймовый монитор может иметь размер зерна 0,24 мм. Каждое зерно состоит из трех точек люминофора, способных светиться красным, зеленым и синим светом соответственно. Наиболее экологичны электронно-лучевые мониторы с повышенной частотой кадровой развертки (75 Гц и более).

Жидкокристаллические дисплеи утомляют зрение меньше, чем электронно-лучевые. Жидкокристаллический дисплей представляет собой две тонкие пластины из кварцевого стекла, на внутренней стороне которых нанесены матрицы из электродов, каждая «ячейка» соответствует одному пикселу. На внешние части стекол наносится поляризующая пленка. Затем две кварцевые панели складываются вместе, между ними создается вакуум и закачивается состав из жидких

кристаллов. Сзади панель подсвечивается источником света. Жидкие кристаллы — органические полимеры, изменяющие прозрачность при приложении электрического поля. В отличие от мониторов на электронно-лучевой трубке, разрешение которых можно менять достаточно гибко, ЖК-дисплеи имеют фиксированный набор «физических» пикселей, поэтому они рассчитаны на работу с максимальным разрешением, заданным производителем. Например, мониторы формата 1024×768 содержат 1024 элемента по горизонтали и 768 — по вертикали. Если пользователь пожелает перевести этот монитор в режим 640×480 , то изображение будет выводиться на середину экрана, а по краям будут оставаться темные поля. Возможно применение интерполяции, тогда изображение займет весь экран, но будет не очень качественным, например очертания букв будут не плавными, а ступенчатыми.

Твердые копии получают с помощью принтеров и плоттеров. Матричные принтеры малопригодны для вывода изображений. Струйные принтеры обеспечивают значительно лучшее качество при той же цене, что и матричные. В конструкции современных принтеров чаще всего используется одна из двух технологий — пьезоэлектрическая или термическая, различающиеся способом выбрасывания струи красителя из печатающей головки.

Пьезоэлектрическая техника drop-on-demand (капля по требованию) использует печатающие головки, в которых размещены миниатюрные пьезоэлектрические пластинки. Под воздействием электрических импульсов с компьютера пластинка изменяет свои размеры, а в заполненном чернилами канале увеличивается давление, что в свою очередь приводит к выбрасыванию капли красителя со скоростью около 20 м/с. Диаметр сопла составляет не менее 15–25 мкм (это ограничено технологией), диаметр вылетевшей из него капли примерно вдвое превышает диаметр сопла (30–50 мкм), при ударе о лист бумаги капля несколько расплывается. Таким образом, разрешение струйного принтера при диаметре сопла 20 мкм составляет 25 мм/0,04 мм, т.е. около 600 точек на дюйм.

В струйных принтерах других конструкций используется термическая разновидность технологии «капля по требованию», т.е. явление увеличения объема тела под влиянием роста температуры и изменения агрегатного состояния. В головке канала, через который проходят чернила, установлены миниатюрные нагревательные элементы, способные за несколько микросекунд достигнуть высокой температуры. При этом создается пузырек пара, выталкивающий каплю красителя из сопла со скоростью до нескольких десятков метров в секунду в то время, как головка «втягивает» следующую порцию чернил.

Современные струйные принтеры позволяют печатать 2–10 страниц формата А4 в минуту с разрешением до 1400 точек на дюйм. Некоторые струйные принтеры реализуют такое же качество изображений, какое достигается на цветных фотографиях небольшого размера. Практически все струйные принтеры позволяют печатать цветные изображения.

Более совершенны лазерные принтеры, использующие технологию Хехо. На специальном фоточувствительном барабане лучом света (лазер, но может применяться и линейка светоизлучающих диодов) создаются заряженные статическим электричеством области, к которым притягиваются частички красящего порошка (тонера). Затем барабан передвигается над листом бумаги, который заряжен еще сильнее. Порошок переносится на бумагу и далее спекается при нагревании, создавая водоупорный отпечаток. Лучшие лазерные принтеры, согласно рекламе, обеспечивают печать 10 и более страниц в минуту при разрешении 2000 точек на дюйм. Наиболее распространены черно-белые лазерные принтеры.

Однако указанное разрешение для струйных и лазерных принтеров реализуется при печати не полутоновых, а бинарных изображений, к числу которых можно отнести буквы текста, черные линии на белом фоне и т.д. Для печати полутоновых изображений, как правило, приходится имитировать полутона с помощью растра. Например, растр полутонового изображения представляет собой ячейки, содержащие 8×8 пикселей; в зависимости от заполнения ячейки точками черного цвета можно имитировать $8^2 = 64$ градации полутонов. Незаполненная ячейка имеет белый цвет, заполненная 8 точками имитирует 12%-ный уровень почернения, заполненная 32 точками — 50%-ный уровень. Так же обстоит дело и при цветной печати красителем каждого цвета. Следовательно, при печати изображений разрешения принтера по полутонам и по элементам зависят друг от друга. Обычно, если разрешение принтера составляет 600 точек на дюйм, то имитируют 64 полутона, используя для этого $600/8 = 75$ ячеек на дюйм. Лучшего результата можно достичь, если отказаться от строгого соответствия полутоновых ячеек и печатных точек, т.е. использовать так называемый диффузионный (стохастический) метод, который получил достаточно широкое распространение.

Для вывода изображений используют также барабанные устройства типа фототелеграфных аппаратов. На барабане закрепляют фотопленку или фотобумагу, сфокусированный луч света переносит изображение на фотоноситель. Такие устройства в настоящее время считаются устаревшими.

2.5. ПРЕДСТАВЛЕНИЕ ИЗОБРАЖЕНИЙ В ЭВМ

При растровом способе записи изображений в памяти ЭВМ формируется машинный кадр в виде совокупности N строк по M пикселей в каждой (см. рис. 2.2). При векторной форме записи в ячейках памяти записывают числа: координаты пиксела x, y и его яркость f (это могут быть и данные о цвете) либо координаты начала и конца прямой линии и ее яркость (цвет). Из линий могут быть образованы сложные фигуры. Векторные данные могут быть представлены также в виде алгоритма. Например, информация о том, что изображение является окружностью цвета C , радиуса R с координатами центра x_0, y_0 , может быть записана в векторной форме на Бейсике как `CIRCLE (x_0, y_0), R , C` .

Пусть изображение представляет собой две прямые с известными координатами начала и конца каждой (рис. 2.4, а).

Это изображение, содержащее две прямые с яркостью w на фоне a в виде 3 строк по 10 пикселей, может быть записано в памяти ЭВМ как `aaaawwwwwwwaaaawaaaaaaaawaaaaa` (рис. 2.4, б). Необходимо привести также данные о числе строк и пикселей в строке.

Векторный файл содержит координаты начала и конца каждой прямой, например $(x_1, y_1; x_2, y_2; x_3, y_3)$. Атрибуты файла могут содержать данные о яркости линий и фона.

Каждая форма записи имеет свои достоинства и недостатки. Растровая форма записи более универсальна, более информативна, идеально соответствует архитектуре ЭВМ. Однако если уменьшить растровое изображение, то часть пикселей сольется. Правда, такое изображение можно записать в память ЭВМ и затем снова увеличить. Получится более грубое изображение, в частности линии станут толще, многие детали исчезнут.

Векторная графика лишена этого недостатка. При любом масштабе точка изображается пикселем, линия имеет одинаковую толщину. Это чрезвычайно важно при работе с геоинформационными системами (ГИС), когда необходимо последовательно переходить от мелкомасштабных изображений к крупномасштабным — от карт местности к планам, содержащим изображения городов, жилищных кварталов и отдельных домов с различными ком-

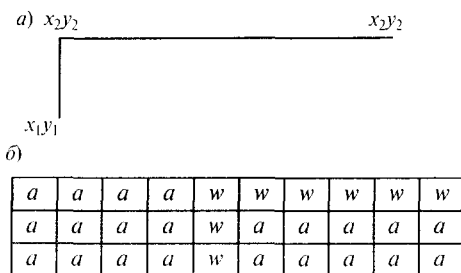


Рис. 2.4. Изображение двух прямых

муникациями, причем изображения сопровождаются текстовой и цифровой информацией. Векторная графика более приспособлена к работе с базами данных, чем растровая. Она в принципе более экономична, чем растровая (нет нужды приводить информацию о «пустых» пикселах) и потому очень эффективна при работе с географическими картами, поскольку в векторной форме в памяти ЭВМ хранятся лишь данные о линиях на карте и их атрибуты. При работе с космическими изображениями Земли приходится одновременно использовать растровую и векторную графику, например накладывать изображение земной поверхности на географическую карту, координатную сетку — на изображение и т.д. Поэтому в программном обеспечении станций приема и обработки космической информации предусматривается преобразование вектор—растр.

Для хранения растровых изображений требуется значительный объем памяти ЭВМ. Если формат кадра составляет 512×512 пикселей при 256 уровнях квантования полутонового изображения, то для его хранения необходимо 250 кбайт, а для цветного (RGB) изображения — 750 Кбайт. При формате 1024×1024 пикселей и 5-канальном изображении, получаемом со сканера AVHRR спутника NOAA (10-битное квантование), требуется 6,25 Мбайт. Не все пиксеты несут информацию, нередко одни и те же значения повторяются много раз, например если какое-либо изображение появляется на однородном фоне. В настоящее время разработаны различные методы экономичной записи изображений в память ЭВМ и сжатия изображений, существует много форматов записи.

2.6. АЛГОРИТМЫ СЖАТИЯ ИЗОБРАЖЕНИЙ

Один из самых простых методов сжатия — групповое кодирование. В соответствии с этой схемой серии повторяющихся величин заменяются одной величиной с указанием количества. Пусть строка изображения имеет вид *abbbcccddeeeeeeeef*. Вместо нее можно записать *1a3b3c2d9e1f*. Этот метод легко реализуем, особенно при работе с длинными сериями повторяющихся величин, например при наличии больших областей с одинаковой яркостью либо цветом. Если серия содержит различные, но незначительно отличающиеся величины, можно указывать и скачок яркости. Групповое кодирование используется в форматах MacPaint, TIF, PCX, BMP.

Цветной вариант рис. 1.2 размером 780×780 пикселей при 24-битном кодировании RGB занимает 1,46 Мбайт. Групповое кодирование в формате BMP сжимает его до 479 кбайт. Опыт показал, что при груп-

повом кодировании легко сжимаются аэрокосмические изображения природных объектов.

В большинстве случаев статистическое кодирование эффективнее, чем групповое. В теории информации доказано, что наибольшим средним количеством информации на одно сообщение (энтропией) обладает источник, выдающий сообщения со случайными, независимыми, равномерно распределенными значениями. Если интервал изменения этих значений равен $[a, b]$, то вероятность появления любого значения из этого интервала одинакова, значения яркости необходимо кодировать словами одинаковой длины $(b - a)$. С этой точки зрения «наиболее информативным изображением» будет изображение, пикселы которого статистически независимы и представляют собой шум с равномерным законом распределения. Пусть изображение содержит N строк, по M пикселов в строке. Для записи такого изображения требуется $NM(b - a)$ бит. В частности, при 1024 строках, 1024 пикселах в строке и $(b - a) = 256$ уровнях квантования изображение занимает 1 Мбайт. Однако реальные изображения существенно отличаются от такого шумового (случайного) двумерного поля (см., например, рис. 2.12). Они содержат фон и некоторые связанные объекты. И фон, и объекты, как правило, содержат повторяющиеся значения яркости, нельзя говорить и о статистической независимости пикселов. Такое изображение обладает *избыточностью*. Если избыточность устранить путем рационального кодирования, учитывающего неравную вероятность яркостей пикселов и их взаимозависимость, то для хранения изображения потребуется меньший объем памяти.

В рамках теории информации в конце 40-х годов XX в. был предложен код Шеннона—Фано, который учитывает неравную вероятность сообщений (букв, цифр, их комбинаций) и использует короткие кодовые слова для наиболее часто появляющихся сообщений, а более длинные — для сообщений, появляющихся редко. Еще более эффективен появившийся чуть позже код Хаффмана, длина кодового слова которого обусловлена вероятностью появления сообщения. При кодировании изображений методом Хаффмана сначала требуется определить вероятность (точнее, частоту) появления яркостей (либо основных цветов) пикселов изображения и с учетом этих вероятностей составить кодовую таблицу. Далее на основании таблицы всем яркостям присваиваются кодовые слова. При этом происходит сжатие изображения, так как для передачи этой совокупности (или записи ее в память ЭВМ) потребуется меньше бит, чем для передачи (записи) кодом с одинаковым числом бит для каждого пиксела.

Пусть растровое изображение имеет вид *abbbcccddeeeeeeeef*. Значения яркости a, b, c, d, e, f появляются с частотами $a:1, b:3, c:3, d:2, e:9$,

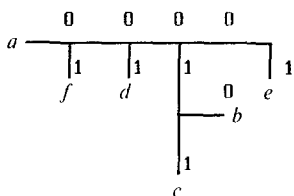


Рис. 2.5. Кодирование по Хаффману

$f:1$. Для кодирования этих шести величин можно использовать равномерный 3-битный код, например: $a: 001$; $b: 011$; $c: 100$; $d: 101$; $e: 110$; $f: 111$. Общая длина кодовой комбинации равна $19 \cdot 3 = 57$ бит.

Для кодирования методом Хаффмана используем двоичное дерево (рис. 2.5). Наиболее редко встречаю-

щиеся в этом примере значения a и f становятся первой парой: a присваивается 0-я ветвь, f — 1-я. 0 и 1 станут младшими битами кодов для a и f . Суммируя частоты a и f , получаем 2. Эта пара объединяется с d , которое также имеет частоту 2. Исходной паре присваивается 0-я ветвь этого дерева, а d присваивается ветвь 1. Теперь код для a заканчивается на 00, для f — на 01, для d — на 1. В процессе построения дерева наименее распространенные величины описываются более длинными кодами, а наиболее распространенные — одним битом (или более). В нашем случае имеем $a: 0000$; $b: 010$; $c: 011$; $d: 001$; $e: 1$; $f: 0001$, общая длина кодовой комбинации равна 41, коэффициент сжатия 1,39:1.

В определенных случаях алгоритм Хаффмана позволяет достигать сжатия больше 8:1. Однако он требует два прохода: один для создания кодовой таблицы, второй — для кодирования. По этой причине и из-за того, что коды переменной длины требуют много времени для компрессии и декомпрессии, кодирование и декодирование методом Хаффмана — сравнительно медленные процессы.

Метод используют, например, при архивировании файлов в ЭВМ, а также в ряде случаев при записи изображений в формате TIF и др. В программном обеспечении современных ЭВМ содержатся модули, позволяющие архивировать и разархивировать файлы, сжимать и восстанавливать сжатые изображения.

Существуют процедуры, не требующие предварительного создания кодовой таблицы, поскольку в них кодовая таблица создается и уточняется по мере поступления данных (*адаптивные методы сжатия*). К числу таких процедур относится схема сжатия LZW (Лемпел—Зива—Велча), которая используется в форматах GIF и TIF.

24-битный цветной вариант изображения на рис. 1.2 размером 780×780 пикселей, 1,46 Мбайт, сжатый по методу LZW в формате TIF, занимает 436 кбайт против 479 кбайт при групповом кодировании в формате BMP.

Рассмотренные методы не охватывают всех известных схем сжатия изображений. Их положительно отличает отсутствие потери информации при сжатии, что важно при обработке и хранении космической

информации. Существуют алгоритмы сжатия с потерями, к числу которых относится JPEG (Joint Photographic Experts Group), расширение JPG. Этот алгоритм использует идею, реализованную в современных системах цветного телевидения (SECAM, PAL, NTSC), и применяется при обработке фотографий и других подобных изображений, а также в издательских системах. При этом учитывается, что человеческий глаз очень чувствителен к изменению яркости изображения, но не замечает цвет мелких его деталей.

Алгоритм JPEG переводит цветное изображение из пространства цвета RGB в пространство YCbCr (как в телевидении): Y определяет яркость пиксела, Cb задает насыщенность для синего цвета, Cr — насыщенность красного. Насыщенность зеленого формируется комбинацией Y, Cb, Cr. Если исключить цветность, т.е. Cb, Cr, то, используя Y, можно получить черно-белое (полутоновое) изображение.

После перехода к пространству YCbCr данные фильтруются отдельно для Y, Cb, Cr в окне 8×8 пикселей с использованием дискретного преобразования Фурье. При этом оцениваются амплитуды гармонических составляющих, те составляющие, амплитуды которых ниже порога, отбрасываются. Для данных Cb и Cr порог устанавливается более высоким, чем для Y, и мелкие детали изображения становятся черно-белыми. Далее Y, Cb, Cr кодируются кодами переменной длины методом Хаффмана.

Алгоритм JPEG обеспечивает очень высокий коэффициент сжатия. Цветной вариант (RGB) изображения на рис. 1.2 размером 780×780 пикселей, 1,46 Мбайт, сжатый по методу JPEG, занимает всего 65 кбайт. JPEG нередко может восстанавливать полноцветное изображение, субъективно неотличимое от подлинника, используя несколько больше одного бита на пиксел при хранении. Алгоритм JPEG не слишком удачен для изображений с небольшим количеством цветов и резкими границами, например для мультипликации, и малопригоден для хранения изображений, получаемых при исследовании Земли из космоса и предназначенных для последующей обработки на ЭВМ.

Некоторая потеря информации происходит при фрактальном сжатии изображений. Понятие фрактала достаточно употребительно в математике и физике. *Фрактальное свойство* плоских изображений состоит в том, что если выбрать одну из связанных точек в качестве центра окружности, то «масса вещества» m , сосредоточенного внутри окружности, будет зависеть от радиуса по закону $m(R) \sim R^D$, где $D < 2$ — параметр фрактальной размерности. Фракталы удачно описывают объекты с «пористой» структурой, основными элементами которой являются фрактальные пакеты или фрактальные кластеры. В изображении отыскиваются фрактальные пакеты, которые затем компактно описываются

ся с помощью математических выражений. В этом смысле фрактальное представление подобно векторной форме записи изображения. Фрактальное сжатие прекрасно сохраняет линии и мелкие детали, но крупные участки с равномерной раскраской приобретают вид мазков кистью. Определенным недостатком фрактального сжатия и восстановления является большой объем вычислений, особенно для сжатия (в десятки раз медленнее, чем по алгоритму JPEG). Фрактальное сжатие относится к числу перспективных методов хранения изображений, полученных при дистанционном зондировании Земли из космоса.

Кроме дискретного преобразования Фурье, при обработке изображений применяют и другие дискретные ортогональные преобразования, в частности преобразование Уолша, используемые в новых — быстрых — алгоритмах сжатия изображений.

Пространства цвета RGB и YCbCr не единственный способ представления цветных изображений в ЭВМ. Для работы с цветными принтерами и другими печатающими устройствами изображение из пространства RGB или YCbCr переводится в CMYK, в котором представлены голубой (C), пурпурный (M), желтый (Y) и черный (K) цвета. При печати используются красители этих цветов, причем черный краситель повышает контраст, тем более, что качественный черный цвет невозможно получить, смешивая голубой, пурпурный и желтый цвета. Введение еще одного способа представления цвета продиктовано особенностями зрения. На дисплее мы видим светящееся изображение в падающем свете, на бумаге — изображение в отраженном свете, глаз регистрирует те цвета, которые остались после поглощения падающего света красителями.

2.7. ФОРМАТЫ ГРАФИЧЕСКИХ ФАЙЛОВ

Каждый графический файл состоит из двух основных частей: заголовка и собственно данных. Цветные изображения содержат также таблицу, в соответствии с которой элементам изображения присваиваются значения основных цветов. Заголовок начинается с идентификатора, указывающего, в каком формате (TIF, BMP, GIF и т.д.) записан файл. Компьютер распознает формат не по расширению, а именно по идентификатору. Тем не менее расширение, по возможности, следует сохранять. После идентификатора приводятся общие сведения о структуре файла (ширина и высота изображения, цветное или полутоновое, использование сжатия и т.д.). Форма представления этих сведений различная для каждого формата.

Наиболее развитым, но и самым сложным является так называемый TIF-формат, файлы которого может читать и записывать каждая серь-

енная программа обработки изображений. В этом формате можно хранить все виды изображений, т.е. монохромные (бинаризованные и полутоновые), цветные с палитрой из 16 и 256 цветов и RGB-изображения с 24-битным кодированием. Наряду с основной информацией об изображении (размеры изображения, данные о цвете) в заголовке TIF-файла можно записать множество дополнительных сведений об изображении.

Из-за сложности TIF-формата не исключено, что некоторые графические редакторы могут создать TIF-файлы с ошибками или не смогут прочитать конкретные TIF-файлы, а некоторые вообще не могут работать со сжатыми TIF-файлами. Однако в любом случае рекомендуется сначала попытаться произвести обмен видеофайлами между различными программами на основе несжатых TIF-файлов, а на другие форматы следует переходить только при наличии сложных проблем.

Формат PCX используется уже в течение многих лет. Несмотря на простоту его структуры, иногда возникают проблемы, особенно в старых программах, поскольку при обмене видеоданными между старыми программами иногда может потеряться информация о разрешении.

Для обмена данными между компьютерами различных систем пригоден еще формат Targa (TGA), не создающий никаких проблем и практически исключающий несовместимость программ. Но и он имеет недостаток: разрешение изображения в файле обычно не запоминается.

Формат GIF известен всем пользователям персонального компьютера, поскольку большое число библиотек изображений записано именно в этом формате. Он обеспечивает максимальное сжатие видеоданных при их записи в память, тем самым уменьшая объем файлов и минимизируя затраты на их загрузку. Стандартная версия формата GIF ограничивается изображениями с палитрой, содержащей максимум 256 цветов (существует версия, позволяющая сохранять RGB-изображение с 8 битами на цвет). Многие программы позволяют вводить в стандартный GIF-формат RGB-изображения с 8 битами на цвет. При этом изображение подвергается внутреннему преобразованию в 256-цветное, о чем пользователю не сообщается. Что произошло, он заметит только при последующей повторной загрузке этого изображения. В формате GIF также не сохраняется разрешение изображений, поэтому его рекомендуется выбирать тогда, когда требуется сохранить цветное изображение в наиболее компактной форме.

В Microsoft Windows и OS/2 используется формат BMP. В последних версиях он был значительно улучшен и теперь представляет солидную основу для обмена данными. Однако здесь также нередко происходит потеря информации о разрешении изображения, поскольку некоторые программы игнорируют соответствующую информацию, со-

держась в заголовке файла. Кроме того, единица разрешения (число пикселей изображения, приходящихся на один метр) неудобна: при пересчете иногда возникают ошибки. Формат BMP рекомендуется применять тогда, когда не удастся осуществить обмен данными в формате TIF.

В рассмотренных форматах использовано сжатие без потери информации. Заметим, что при этом иногда возможно даже увеличение размера файла: когда записанная информация, необходимая для интерпретации сжатых данных, занимает больше места, чем сэкономлено.

Формат JPEG, как указывалось, предусматривает сжатие с потерей информации.

Кроме рассмотренных существует много других графических форматов. В качестве примера приведем структуру файла в формате BMP, в котором записываются несжатые изображения или сжатые с помощью группового кодирования. В начале файла помещается идентификатор — 4D42h или BM в ASCII. Далее в 14-байтном заголовке стоит число, указывающее размер файла в байтах (если применено сжатие); в несжатых файлах это обычно 0. Поля Reserved 1 и Reserved 2 не содержат полезных данных и обычно устанавливаются в 0. Эти поля часто используются приложениями после считывания заголовка в память. В поле BitmapOffset хранятся сведения о начальном смещении растровых данных от начала файла (в байтах).

В файлах BMP версии 3.x за заголовком файла следует второй заголовок размером 40 байтов, называемый заголовком раstra и содержащий одиннадцать полей:

- 1) Size — размер заголовка в байтах (всегда 40);
- 2) Width — ширина изображения в пикселах;
- 3) Height — высота изображения в пикселах; если Height — положительное число, то изображение представляет собой «восходящий» растр с началом в левом нижнем углу. Если Height — число отрицательное, то изображение представляет собой «нисходящий» растр с началом в левом верхнем углу;
- 4) Planes — число цветовых плоскостей, используемых для представления растровых данных. Файлы BMP всегда состоят из одной цветовой плоскости, поэтому значение этого поля всегда равно 1;
- 5) BitsPerPixel — число бит на пиксел в каждой плоскости. Диапазон его значений 1–24. Для BMP 3.x допустимы только значения 1, 4, 8 и 24;
- 6) Compression — метод кодирования растровых данных. Значение 0 указывает на то, что данные не сжаты, 1 — применен 8-битный алгоритм группового кодирования; 2 — использован 4-битный алгоритм;

7) **SizeOfBitmap** — размер растра в байтах. Значение этого поля равно 0, если растр не сжат (в этом случае декодер определяет изображение по его размерам);

8–9) **HorzResolution** и **VertResolution** — соответственно горизонтальное и вертикальное разрешения, выраженные в пикселах на метр. Эти поля помогают программе чтения BMP выбирать необходимое разрешение при печати и воспроизведении BMP-файла;

10) **ColorsUsed** — число цветов в палитре. Если значением этого поля является 0, а значение **BitsPerPixel** меньше 16, то число элементов равно максимально возможному для этой цветовой таблицы. Если значение поля **BitsPerPixel** больше или равно 16, то цветовая палитра отсутствует, значение поля **ColorsUsed** вычисляется из значения поля **BitsPerPixel**: $\text{ColorsUsed} = \text{BitsPerPixel}$;

11) **Colorsimportant** — число наиболее важных цветов в палитре, определяемое частотой их появления в растровых данных. Это поле используется для обеспечения максимально возможного качества изображения на графических устройствах, которые не поддерживают необходимое количество цветов. Пусть в 8-битном изображении, содержащем, например, 142 цвета, основную часть изображения составляют всего 10–20 цветов. Если эти цвета будут определены, то 16-цветный адаптер дисплея отобразит данное изображение достаточно точно, используя для этого 16 наиболее часто встречающихся цветов. Самые важные цвета в палитре всегда хранятся первыми. Если все цвета в цветовой таблице рассматриваются как важные, то значение поля **Colorsimportant** равно 0.

После заголовка растра может следовать цветовая палитра. В полях **Blue**, **Green** и **Red** содержатся цветовые составляющие для пиксела (значение каждой может находиться в диапазоне 0–255). Поле **Reserved** служит для заполнения структуры до границы двойного слова и всегда равно 0.

Далее помещаются биты изображения, которые раскодируются компьютером с использованием данных, записанных в заголовках.

2.8. МОДЕЛИ РАСТРОВЫХ ИЗОБРАЖЕНИЙ

При исследовании алгоритмов обработки используются различные математические модели изображений. Линейное изображение, как уже указано, можно записать в векторном виде, например как совокупность ломаных линий. Сложнее обстоит дело с растровыми изображениями.

Для обнаружения (выделения) конкретных объектов на изображении обычно применяют детерминированные модели, например запол-

ненный контур стилизованного изображения самолета, корабля или какой-либо детали на конвейере. Детерминированным объектом является географическая карта или ее элементы, например гидросеть, используемые при отождествлении космо- и аэроизображений земной поверхности.

Растровые изображения земной поверхности и облаков, получаемые при дистанционном зондировании Земли из космоса, как правило, представляют собой совокупность пятен неправильной, случайной формы, имеющих различную яркость и цвет (см. рис. 1.1, 1.16 и 2.2). Для математического описания такого рода плоских изображений и создания их моделей, а также шума и помех на изображении применяются двумерные случайные поля.

2.8.1. Двумерное случайное поле

Двумерное случайное поле — это последовательность отсчетов, в нашем случае отсчетов яркости пикселей $\{f\}$, каждый из которых представляет собой случайную величину с некоторым распределением вероятностей. Пусть x_i — координата пикселя вдоль строки, y_j — координата вдоль столбца, \mathbf{r}_{ij} — радиус-вектор точки с координатами (x_i, y_j) , i и j — номера строки и столбца, $i = 1, 2, \dots, N$; $j = 1, 2, \dots, M$, N и M — число строк и столбцов, $P = N \cdot M$ — общее число пикселей изображения.

Будем считать для простоты, что яркость f — непрерывная случайная величина. Обозначим через $w_1[f(\mathbf{r}_{ij})]$ одномерную плотность вероятности яркости пикселя с координатами, определяемыми радиусом-вектором \mathbf{r}_{ij} . Произведение $w_1[f(\mathbf{r}_{ij})]df$ — это вероятность попадания яркости в интервал значений от $f(\mathbf{r}_{ij})$ до $f(\mathbf{r}_{ij}) + df$. Вероятность попадания яркости в весь интервал возможных значений от A до B равна 1 (условие нормировки):

$$\int_B^A w_1[f(\mathbf{r}_{ij})]df = 1. \quad (2.1)$$

Математическое ожидание случайной величины $f(\mathbf{r}_{ij})$ определяет среднюю яркость пикселей изображения:

$$\mu_1(\mathbf{r}_{ij}) = \langle f(\mathbf{r}_{ij}) \rangle = \int_B^A f(\mathbf{r}_{ij}) w_1[f(\mathbf{r}_{ij})]df. \quad (2.2)$$

Дисперсия непрерывной случайной величины $f(\mathbf{r}_{ij})$ характеризует разброс значений яркости относительно математического ожидания:

$$\sigma(\mathbf{r}_{ij}) = \left\langle [f(\mathbf{r}_{ij}) - \mu(\mathbf{r}_{ij})]^2 \right\rangle = \int_B [f(\mathbf{r}_{ij}) - \mu(\mathbf{r}_{ij})] w_i |f(\mathbf{r}_{ij})| df. \quad (2.3)$$

В этих формулах для краткости символом $\langle \rangle$ обозначена операция усреднения по ансамблю реализаций случайного поля $\{f\}$ с использованием плотности вероятности.

Кроме одномерной плотности вероятности в практике цифровой обработки изображений приходится иметь дело с плотностями вероятности более высоких порядков. Обычное RGB-изображение — это совокупность значений яркости в трех спектральных каналах, спутниковые многоспектральные изображения — совокупность значений яркости в двух и более спектральных каналах. Рассматривая яркости одного пиксела в нескольких спектральных каналах как вектор в соответствующем пространстве, можно говорить о совместном (многомерном) законе распределения компонентов вектора. Наиболее полное статистическое описание изображения дает совместный закон распределения всех векторов изображения.

Двумерная плотность вероятности $w_{11}[f_a(\mathbf{r}_{ij}), f_b(\mathbf{r}_{ij})]$ характеризует, в среднем, совместные статистические свойства двух компонентов $f_a(\mathbf{r}_{ij})$ и $f_b(\mathbf{r}_{ij})$ вектора $\mathbf{f}(\mathbf{r}_{ij})$ с координатами x_i и y_j . Произведение $w_{11}[f_a(\mathbf{r}_{ij}), f_b(\mathbf{r}_{ij})] \cdot df_a(\mathbf{r}_{ij}) df_b(\mathbf{r}_{ij})$ равно вероятности попадания указанных компонентов в прямоугольник со сторонами $[f_a(\mathbf{r}_{ij}), f_a(\mathbf{r}_{ij}) + df_a(\mathbf{r}_{ij})]$ и $[f_b(\mathbf{r}_{ij}), f_b(\mathbf{r}_{ij}) + df_b(\mathbf{r}_{ij})]$.

Двумерная плотность вероятности $w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})]$ характеризует, в среднем, совместные статистические свойства двух векторов $\mathbf{f}(\mathbf{r}_{ij})$ и $\mathbf{f}(\mathbf{r}_{kl})$ с координатами x_i, y_j и x_k, y_l соответственно. Произведение $w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})] \cdot d\mathbf{f}(\mathbf{r}_{ij}) d\mathbf{f}(\mathbf{r}_{kl})$ равно вероятности попадания указанных векторов в прямоугольник со сторонами $[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{ij}) + d\mathbf{f}(\mathbf{r}_{ij})]$ и $[\mathbf{f}(\mathbf{r}_{kl}), \mathbf{f}(\mathbf{r}_{kl}) + d\mathbf{f}(\mathbf{r}_{kl})]$.

Двумерная плотность вероятности, как и одномерная, нормируется к 1. Проинтегрировав ее по всему интервалу возможных значений яркости одной из переменных, например $\mathbf{f}(\mathbf{r}_{ij})$, получим одномерную плотность вероятности $w_1[\mathbf{f}(\mathbf{r}_{kl})]$.

Если два вектора (два компонента одного вектора) статистически независимы (точнее, некоррелированы), то двумерная плотность вероятности распадается на произведение одномерных:

$$\begin{aligned}w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})] &= w_1[\mathbf{f}(\mathbf{r}_{ij})] \cdot w_1[\mathbf{f}(\mathbf{r}_{kl})], \\w_{11}[f_a(\mathbf{r}_{ij}), f_b(\mathbf{r}_{ij})] &= w_1[f_a(\mathbf{r}_{ij})] \cdot w_1[f_b(\mathbf{r}_{ij})].\end{aligned}\quad (2.4)$$

Важные свойства случайного поля описывают их корреляционные характеристики. Коэффициент корреляции двух компонент a и b вектора $\mathbf{f}(\mathbf{r}_{ij})$ определяется как

$$R_{ab}(\mathbf{r}_{ij}) = \int_A \int_B (f_a - \mu_{a1})(f_b - \mu_{b1}) w_{11}(f_a, f_b) df_a df_b. \quad (2.5)$$

Здесь $f_a = f_a(\mathbf{r}_{ij})$ и $\mu_{a1} = \mu_{a1}(\mathbf{r}_{ij})$ — яркость и математическое ожидание яркости компонента a вектора $\mathbf{f}(\mathbf{r}_{ij})$, f_b и μ_{b1} — то же для компонента b .

Пусть $\sigma_a = \sigma_a(\mathbf{r}_{ij})$ — среднеквадратическое отклонение (СКО) яркости для компонента a ; $\sigma_b = \sigma_b(\mathbf{r}_{ij})$ — СКО компонента b . Отношение $R_{ab}(\mathbf{r}_{ij})/\sigma_a\sigma_b = \rho_{ab}(\mathbf{r}_{ij})$ носит название *нормированного коэффициента корреляции* компонент вектора $\mathbf{f}(\mathbf{r}_{ij})$ для пиксела с координатами x_i, y_j ; $-1 \leq \rho_{ab}(\mathbf{r}_{ij}) \leq 1$. Если компоненты вектора $\mathbf{f}(\mathbf{r}_{ij})$ статистически независимы, то $\rho_{ab}(\mathbf{r}_{ij}) = 0$.

При дистанционных исследованиях значение $\rho_{ab}(\mathbf{r}_{ij})$ определяется зависимостью отражательной способности наблюдаемого объекта от длины волны. Например, для облаков яркость в первом a (видимом) канале сканера AVHRR спутника NOAA приблизительно равна яркости во втором b (ближнем ИК) канале, поэтому $\rho_{ab}(\mathbf{r}_{ij}) \approx 1$; снег во втором канале AVHRR выглядит приблизительно в 3–5 раз темнее, чем в первом, и для него $\rho_{ab}(\mathbf{r}_{ij}) \approx 0,2-0,3$.

Функция автокорреляции двух векторов $\mathbf{f}(\mathbf{r}_{ij}) = \mathbf{f}_{ij}$ и $\mathbf{f}(\mathbf{r}_{kl}) = \mathbf{f}_{kl}$ с математическими ожиданиями $\mu_{1ij} = \mu_1(\mathbf{r}_{ij})$ и $\mu_{1kl} = \mu_1(\mathbf{r}_{kl})$ зависит от координат (x_i, y_j) и (x_k, y_l) :

$$\begin{aligned}R_{ab}(\mathbf{r}_{ij}, \mathbf{r}_{kl}) &= \langle (\mathbf{f}_{ij} - \mu_{1ij})(\mathbf{f}_{kl} - \mu_{1kl}) \rangle = \\&= \int_A \int_B (\mathbf{f}_{ij} - \mu_{1ij})(\mathbf{f}_{kl} - \mu_{1kl}) w_{11}(\mathbf{f}_{ij}, \mathbf{f}_{kl}) d\mathbf{f}_{ij} d\mathbf{f}_{kl}.\end{aligned}\quad (2.6)$$

2.8.2. Однородные и изотропные случайные поля

Случайное поле называется однородным по x , если функции $w_1[\mathbf{f}(\mathbf{r}_{ij})]$ и $w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})]$ инвариантны относительно любого конечного сдвига

группы точек $(x_1, \dots, x_i, \dots, x_N)$ — координаты пикселей — вдоль оси Ox (на практике такой сдвиг должен быть небольшим по сравнению с размером кадра). Однородность является обобщением на случайные поля понятия стационарности, введенного для случайных процессов. Случайное поле может быть однородно по одной координате и неоднородно — по другой (рис. 2.6).



Рис. 2.6. Случайное поле однородно по горизонтали и неоднородно по вертикали

Очевидно, что статистические характеристики случайного поля, описывающего это изображение, например плотность вероятности $w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})]$, не изменятся, если в пределах кадра изменить начало отсчета по оси Ox . В то же время изменение начала отсчета по оси Oy приведет к изменению статистических характеристик.

Одномерная плотность вероятности однородного по всем координатам случайного поля $w_1[\mathbf{f}(\mathbf{r}_{ij})]$, по определению, инвариантна относительно изменения начала отсчета координат. Поместив начало отсчета в точку (x_i, y_j) , находим, что w_1 не должна зависеть от координат, т.е. $w_1 = w_1(f)$. Соответственно не зависят от координат математическое ожидание яркости пикселей μ_1 и дисперсия σ^2 , постоянные для всего поля.

Приведем два примера одномерной плотности вероятности $w_1(f)$. Нормальная (гауссовская) плотность вероятности (рис. 2.7, а) описывается выражением

$$w_1(f) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(f - \mu_1)^2 / 2\sigma^2\right], \quad (2.7)$$

где μ_1 — среднее значение яркости, σ^2 — дисперсия.

Для равномерной плотности вероятности (рис. 2.7, б) все значения яркости, от минимальной до максимальной, равновероятны.

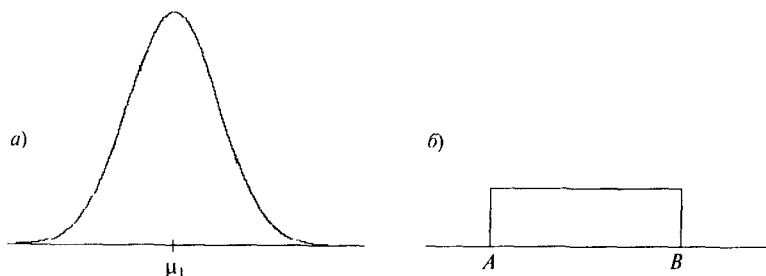


Рис. 2.7. Нормальная (а) и равномерная (б) плотности вероятности

Обе эти величины часто встречаются в задачах обработки изображений. Однако нормальная плотность вероятности обязательно предполагает, что некоторые значения яркости могут быть отрицательными, хотя яркость всегда положительна. Поэтому нормальная кривая дает приближенное описание плотности вероятности яркости изображения.

Плотность вероятности двух векторов однородного по всем координатам случайного поля $w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})]$ зависит от разности радиусов-векторов точек с координатами (x_i, y_j) и (x_k, y_l) , т.е. от $(\mathbf{r}_{ij} - \mathbf{r}_{kl})$, а функция автокорреляции двух векторов имеет вид $R(\mathbf{r}_{ij}, \mathbf{r}_{kl}) = R(\mathbf{r}_{ij} - \mathbf{r}_{kl})$.

Если плотность вероятности однородного по всем координатам случайного поля $w_{11}[\mathbf{f}(\mathbf{r}_{ij}), \mathbf{f}(\mathbf{r}_{kl})]$ инвариантна относительно произвольного поворота, то поле называется однородным и изотропным, в этом случае w_{11} зависит только от $|\mathbf{r}_{ij} - \mathbf{r}_{kl}|$. Функция автокорреляции R однородного и изотропного поля также зависит только от $|\mathbf{r}_{ij} - \mathbf{r}_{kl}|$.

Для неизотропного поля w_{11} зависит как от модуля вектора $|\mathbf{r}_{ij} - \mathbf{r}_{kl}|$, так и от его направления. Изображение, описываемое однородным, но не изотропным полем, можно видеть на рис. 2.6.

Двумерная плотность вероятности нормального закона распределения яркости одного и того же компонента f двух векторов $\mathbf{f}(\mathbf{r}_{ij})$ и $\mathbf{f}(\mathbf{r}_{kl})$ однородного и изотропного случайного поля с математическим ожиданием μ_1 , дисперсией σ^2 и нормированной функцией автокорреляции $\rho = \rho(|\mathbf{r}_{ij} - \mathbf{r}_{kl}|) = R/\sigma^2$ описывается выражением (2.8), где $f_1 = f(\mathbf{r}_{ij})$ и $f_2 = f(\mathbf{r}_{kl})$:

$$w_{11}(f_1, f_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(f_1 - \mu_1)^2}{\sigma^2} + \frac{(f_2 - \mu_1)^2}{\sigma^2} - \right. \right. \\ \left. \left. - 2\rho \frac{(f_1 - \mu_1)(f_2 - \mu_1)}{\sigma^2} \right] \right\}. \quad (2.8)$$

Для изображения, описываемого однородным и изотропным полем, оценку $w_1(f)$ каждого компонента f вектора \mathbf{f} получают из гистограммы яркости пикселей. Весь диапазон возможных значений яркости компонента разбивается на ряд интервалов, гистограмма показывает, сколько пикселей попало в каждый интервал.

На рис. 2.8 по оси абсцисс отложена яркость (от 0 до 255), по оси ординат — упомянутое число пикселей. Если все ординаты разделить на общее число пикселей $P = N \cdot M$, то получится оценка $w_1(f)$. Гистограмма яркости часто применяется при анализе изображений.

Оценкой математического ожидания компонента f вектора однородного поля является выборочное среднее

$$\bar{\mu}_1 = \frac{1}{P} \sum_{i=1}^N \sum_{j=1}^M f_{ij}, \quad (2.9)$$

оценкой дисперсии — выборочная дисперсия

$$\bar{\sigma}^2 = \frac{1}{P-1} \sum_{i=1}^N \sum_{j=1}^M (f_{ij} - \bar{\mu}_1)^2. \quad (2.10)$$

Функцию автокорреляции некоторого компонента f вектора однородного и изотропного поля можно оценить по формуле

$$\bar{R}_{nm} = \frac{1}{(N-n)(M-m)} \sum_{i=1}^{N-n} \sum_{j=1}^{M-m} [f(\mathbf{r}_{ij}) - \bar{\mu}_1][f(\mathbf{r}_{kl}) - \bar{\mu}_1]. \quad (2.11)$$

Здесь $k = i + n$, $l = j + m$, пиксел $f(\mathbf{r}_{kl})$ смещен относительно $f(\mathbf{r}_{ij})$ на n пикселей по оси $0x$ и на m пикселей по оси $0y$. Индекс nm означает, что расстояние между пикселями $f(\mathbf{r}_{ij})$ и $f(\mathbf{r}_{kl})$ равно $|\mathbf{r}_{ij} - \mathbf{r}_{kl}| = \sqrt{n^2 + m^2}$. Пример функции автокорреляции некоторого компонента f вектора однородного и изотропного поля показан на рис. 2.9.



Рис. 2.8. Гистограмма яркости пикселей изображения

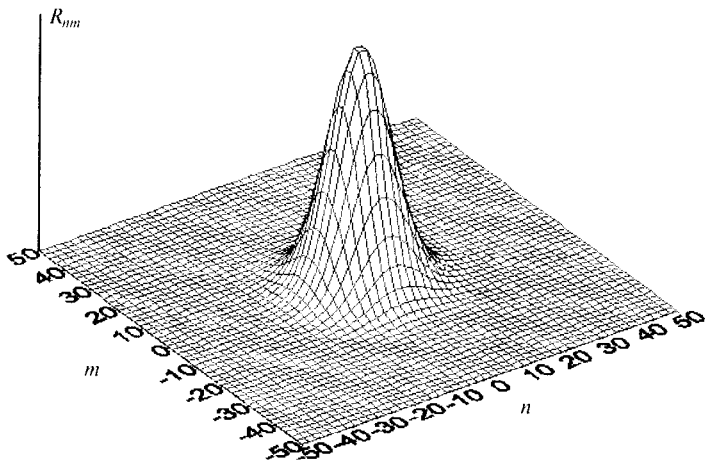


Рис. 2.9. Функция автокорреляции однородного и изотропного случайного поля

Оценка коэффициента корреляции двух компонент a и b вектора \mathbf{f}_{ij} определяется как

$$\bar{R}_{mn} = \frac{1}{P-1} \sum_{i=1}^N \sum_{j=1}^M [f_a(\mathbf{r}_{ij}) - \bar{\mu}_a] [f_b(\mathbf{r}_{ij}) - \bar{\mu}_b].$$

Оценка нормированного коэффициента корреляции компонент a и b вектора \mathbf{f}_{ij} равна $\bar{\rho}_{ab} = \bar{R}_{ab} / \bar{\sigma}_a \bar{\sigma}_b$. Совокупность коэффициентов корреляции всех компонент вектора образует *корреляционную матрицу*.

Математическое ожидание яркости пикселей, дисперсия яркости и их оценки имеют вполне определенный физический смысл. Величина μ_1 характеризует среднюю яркость изображения, среднеквадратическое отклонение σ — контраст, т.е. разницу между яркостями наиболее темных и наиболее светлых участков изображения. Если уменьшить среднюю яркость, то гистограмма на рис. 2.8 сместится влево, если увеличить — вправо. Если увеличить контраст, то гистограмма станет шире, если уменьшить — то уже. О потенциальных возможностях улучшения яркости и контраста судят по гистограмме, и не случайно преобразования яркости и контраста в задачах обработки изображений называют гистограммными (или модификацией гистограмм). Гистограммные преобразования используются не только при обработке изображений со случайной структурой, но и при обработке детерминированных растровых изображений, имеющих вполне регулярную структуру.

По функции автокорреляции изображения, описываемого однородным и изотропным случайным полем, можно определить, на каком

расстоянии r_0 между пикселями статистическая зависимость значений яркости в среднем ослабевает. Это расстояние может задавать характерный размер объектов («пятен») на таком изображении.

Кроме математического ожидания, дисперсии, функции автокорреляции и их оценок в практике обработки изображений используется медиана закона распределения яркости (медианное значение яркости). Вероятность того, что яркость всех пикселей изображения превышает медианное значение, равна 0,5, вероятность того, что эта яркость ниже медианы, также равна 0,5. Для симметричных относительно математического ожидания плотностей вероятности, например нормальной и равномерной, медиана совпадает с μ . Оценку медианы можно найти по гистограмме. Выборочная медиана на рис. 2.8 делит гистограмму пополам: одинаковое число пикселей попадает в участки слева и справа от медианы.

2.8.3. Модели случайных полей

При моделировании случайных полей на ЭВМ числа от датчика случайных чисел преобразуются в числовую матрицу, обладающую определенными свойствами. В зависимости от того, какие отсчеты датчика

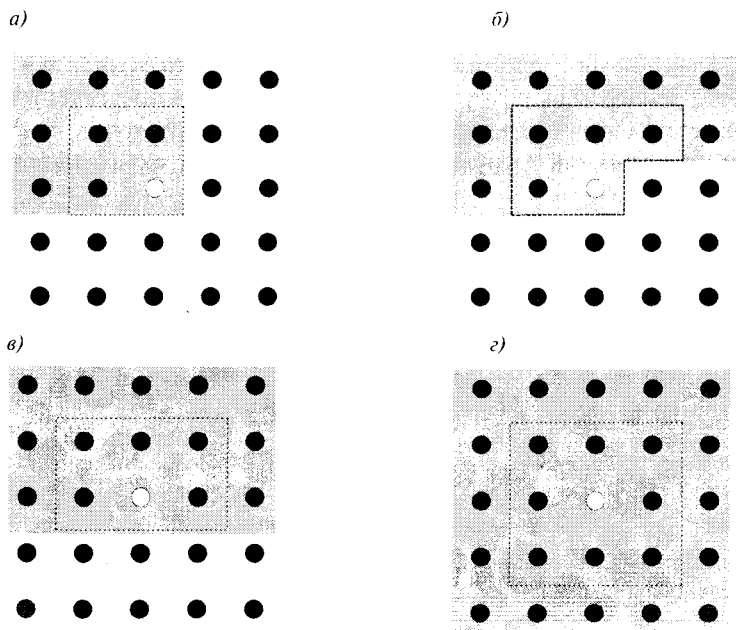


Рис. 2.10. Каузальная, полумаузальная и некаузальная модели

участвуют в формировании текущего значения поля, модели случайных полей разделяются на каузальные, полукаузальные и некаузальные. Основное различие между этими моделями обусловлено их пространственными особенностями, которые зачастую приводят к принципиально различным алгоритмам. Если для формирования отсчета поля (белая точка на рис. 2.10) с координатами (x, y) используются текущий отсчет датчика и предыдущие его отсчеты $x_1 \leq x$ и $y_1 \leq y$ (рис. 2.10, а), то модель называют каузальной (от лат. *causa* — причина).

Иногда под каузальной понимают модель, использующую данные области, геометрия которой определяется растровой разверткой (рис. 2.10, б). При работе в реальном времени, когда известны лишь эти данные и потому выполняется условие причинности, такая схема единственно возможная. Для полукаузального фильтра задействованы входные данные верхней полуплоскости (рис. 2.10, в), а для некаузального — данные всей плоскости (рис. 2.10, г). Могут использоваться не все данные соответствующей области, а лишь некоторая их часть (например, ограниченная пунктирной линией на рис. 2.10).

Пример изображения, построенного по каузальной методике, можно видеть на рис. 2.11. Это модель аэрофотоснимка леса в виде совокупности кружков разного диаметра, случайным образом размещенных на плоскости, причем кружки не перекрываются (в природе имеет место «расталкивание» крон деревьев в спелом древостое). Первоначально по случайному закону выбирается центр первой окружности в окрестности левого верхнего угла, рисуется эта окружность. Далее в ее окрестности случайным образом ставятся следующие точки — центры других окружностей. Окружность воспроизводится, если выполняется

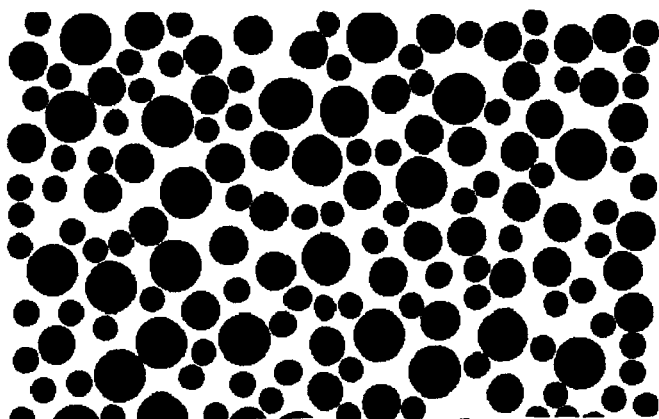


Рис. 2.11. Модель аэрофотоизображения леса

условие $(R_i + R_j)^2 > (x_i - x_j)^2 + (y_i - y_j)^2$, где (x_i, y_i) , (x_j, y_j) — координаты центров ближайших соседей, а R_i , R_j — их радиусы ($i \neq j$). По этой схеме постепенно, сверху вниз заполняется все поле.

В программном обеспечении современных ЭВМ есть алгоритм моделирования случайных чисел, распределенных равномерно в интервале 0—1 и статистически независимых в достаточно длинной последовательности. Математическое ожидание таких чисел $\mu_1 = 0,5$, дисперсия $\sigma^2 = 1/12$. Если вычесть из этих чисел μ_1 , получим последовательность, распределенную равномерно в интервале от $-0,5$ до $0,5$. Из этой последовательности можно получить последовательность нормально распределенных случайных чисел. Из теории вероятностей известно, что сумма независимых одинаково распределенных случайных величин имеет нормальный закон распределения при условии, что слагаемые дают равномерно малый вклад в сумму (центральная предельная теорема). Чтобы получить один отсчет h_k с нормальным законом распределения, достаточно просуммировать 8—10 равномерно распределенных случайных чисел, полученных по описанной методике, однако удобнее использовать 12 чисел.

Пусть исходная реализация $\{x_i\}$ с равномерным законом распределения имеет $\mu_1 = 0$ и $\sigma^2 = 1/12$. Учитывая, что значения этой реализации x_i и x_j при $i \neq j$ статистически независимы, получаем

$$\langle h_k \rangle = 0, \quad \langle h_k^2 \rangle = \sum_{i=1}^{12} \langle x_i^2 \rangle = 12/12 = 1,$$

т.е. реализация $\{h_k\}$ с нормальным законом распределения имеет нулевое среднее и единичную дисперсию. Умножая все h_k на постоянное число σ и прибавляя к ним постоянную μ_1 , получаем реализацию с заданными дисперсией и средним значением.

Используя случайные числа x_i или h_k , можно моделировать изображение, если последовательно, строка за строкой, присваивать уровням яркости пикселей f_j эти случайные значения. Необходимо учитывать, что яркость — величина неотрицательная. Если значения яркости лежат в интервале 0—255, то при нормальном законе распределения можно положить, например, $\mu_1 = 127$, $\sigma = 50$, при этом $f_j < 0$ и $f_j > 255$ маловероятны. Можно задавать большие σ , в этих случаях отрицательным f_j следует присваивать нулевые значения, а $f_j > 255$ присваивать значения $f_j = 255$.

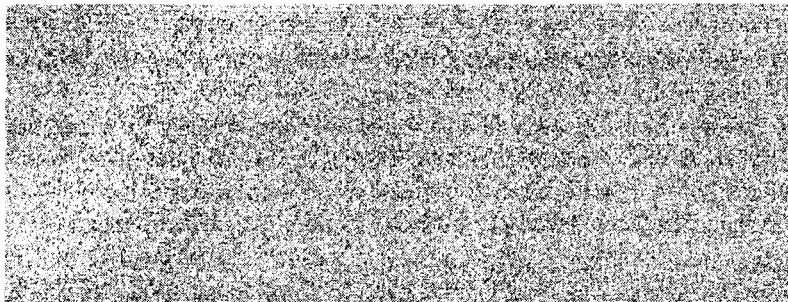


Рис. 2.12. Яркости пикселей статистически независимы

Изображение на рис. 2.12 получено в результате такого моделирования. Использованы независимые случайные значения с нормальным законом распределения и $\mu_1 = 127$, $\sigma = 50$. Поскольку между яркостями пикселей нет связи, то на рис. 2.12 изображен скорее шум, чем земная поверхность.

Более реалистичные изображения удастся получить, если использовать авторегрессионную модель, позволяющую генерировать однородное, однородное и изотропное или неоднородное случайное поле с заданными корреляционной функцией и математическим ожиданием. Авторегрессионные методики типично каузальные.

Рассмотрим простейшую линейную авторегрессионную модель изображения — авторегрессию по строкам (рис. 2.13, а), например модель строки телевизионного сигнала, значения яркости которой считаются вдоль строки с некоторым шагом:

$$f_i = \rho f_{i-1} + (1 - \rho)\mu_1 + \sigma\sqrt{1 - \rho^2}h_i, \quad (2.12)$$

где f_i и f_{i-1} — значения яркости на i -м и $(i-1)$ -м шаге, h_i — одинаково распределенные независимые случайные числа с нулевым средним и единичной дисперсией (порождающий случайный процесс), параметры ρ , μ_1 , σ — постоянные. Отсюда следует, что статистические харак-

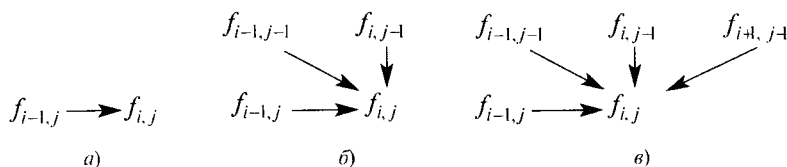


Рис. 2.13. Схемы авторегрессии

теристики случайной последовательности $\{f_i\}$ должны быть инвариантны относительно изменения начала отсчета пикселей в строке, т.е. строка $\{f_i\}$ обладает свойствами однородного случайного поля. Рекуррентное соотношение (2.12) описывает линейное инерционное преобразование случайной последовательности $\{h_i\}$. Усредним правую и левую части выражения: $\langle f_i \rangle = \rho \langle f_{i-1} \rangle + (1 - \rho)\mu_1$, откуда $\langle f_i \rangle = \langle f_{i-1} \rangle = \mu_1$.

Перепишем авторегрессионное уравнение в виде

$$f_i - \mu = \rho(f_{i-1} - \mu_1) + \sigma\sqrt{1 - \rho^2}h_i.$$

Умножим правую и левую часть на $(f_{i-1} - \mu_1)$ и усредним:

$$\langle (f_i - \mu_1)(f_{i-1} - \mu_1) \rangle = \rho \langle (f_{i-1} - \mu_1)^2 \rangle.$$

Среднее от произведения $f_{i-1} h_i$ равно нулю, так как f_{i-1} и h_i статистически независимы (случайное число h_i появилось позже, чем f_{i-1}). Выражение $\langle (f_i - \mu_1)(f_{i-1} - \mu_1) \rangle = R_1$ — это коэффициент корреляции между яркостями двух соседних пикселей в строке.

Возведем левую и правую часть авторегрессионного уравнения в квадрат и усредним:

$$\langle (f_i - \mu_1)^2 \rangle = \rho^2 \langle (f_{i-1} - \mu_1)^2 \rangle + \sigma^2(1 - \rho^2) \langle h_i^2 \rangle.$$

Так как $\langle h_i^2 \rangle = 1$, то $\langle (f_i - \mu_1)^2 \rangle = \langle (f_{i-1} - \mu_1)^2 \rangle = \sigma^2$ — дисперсия случайной последовательности $\{f_i\}$. Параметр $\rho = R_1/\sigma^2$ — это нормированный к σ^2 коэффициент корреляции между f_i и f_{i-1} , $0 \leq \rho \leq 1$.

Умножим левую и правую части авторегрессионного уравнения (2.12) на $(f_{i-2} - \mu_1)$ и усредним. В результате получим $\langle (f_i - \mu_1)(f_{i-2} - \mu_1) \rangle = R_2$ — коэффициент корреляции между значениями яркости двух пикселей, отстоящих друг от друга в строке через один шаг. Можно показать, что $R_2 = \rho^2 \sigma^2$. Продолжая эту процедуру, найдем, что коэффициент корреляции между значениями яркости пикселей, разделенных вдоль строки n шагами, $R_n = \rho^n \sigma^2 = \sigma^2 e^{-\alpha n}$, где $\alpha = \ln(1/\rho)$. Таким образом, случайная последовательность $\{f_i\}$ имеет экспоненциальную функцию автокорреляции; f_i обладает нормальным законом распределения, если h_i распределено по нормальному закону или если закон распределения h_i отличается от нормального, но ρ близко к 1 (имеет место нормализация случайного процесса f_i). Попытка трансформировать

закон распределения путем нелинейного преобразования $\{f_i\}$ приводит к «неуправляемой» функции корреляции, попытка выполнить сначала нелинейное преобразование последовательности независимых отсчетов $\{h_i\}$, а затем линейное не всегда приводит к желаемому результату, поскольку линейная система нормализует случайный процесс при достаточно большом p .

Из рис. 2.14 видно, что существует некоторая зависимость между яркостью пикселей в строке, но каждая соседняя строка случайным образом следует за предыдущей, изображение, скорее, похоже на узор на ткани, чем на снимок земной поверхности, полученный из космоса.

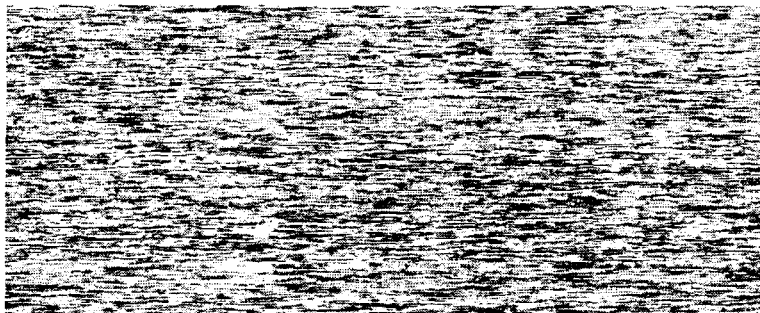


Рис. 2.14. Авторегрессия по строкам

Трехточечная авторегрессионная модель изображения (см. рис. 2.13, б) может быть записана в виде

$$f_{ij} = \mu_1 + \rho_1(f_{i-1,j} - \mu_1) + \rho_2(f_{i,j-1} - \mu_1) - \rho_1\rho_2(f_{i-1,j-1} - \mu_1) + \sigma\sqrt{1 - \rho_1^2 - \rho_2^2 + \rho_1^2\rho_2^2} h_{ij}. \quad (2.13)$$

Здесь i и j — номера строки и пиксела в строке, ρ_1 и ρ_2 — коэффициенты корреляции значений яркости рядом расположенных пикселей в соседних строках и в одной строке. По-прежнему $h_{i,j}$ — одинаково распределенные независимые случайные величины с нулевым средним и единичной дисперсией. Функция автокорреляции значений яркости пикселей, расположенных в одном столбце, но через m строк, имеет вид $R_{m0} = \rho_1^m \sigma^2$, функция автокорреляции яркости пикселей, расположенных в одной строке через n шагов, как и для схемы авторегрессии по строкам, $R_{0n} = \rho_2^n \sigma^2$. Функция автокорреляции яркостей пикселей, разделенных n столбцами и m строками, есть произведение двух экспоненци-

$$R_{mn} = \rho_1^m \rho_2^n \sigma^2 = \sigma^2 e^{-\alpha m} e^{-\beta n}, \text{ где } \alpha = \ln(1/\rho_1), \beta = \ln(1/\rho_2).$$

При $\rho_1 \neq \rho_2$ имеем случайное поле, однородное по строкам и по столбцам, но не изотропное. Можно положить $\rho_1 = \rho_2$, однако при этом случайное поле будет лишь приближенно изотропным. Если представить $\rho_{mn} = R_{mn}/\sigma^2$ в виде поверхности (рис. 2.15), то окажется, что сечение ρ_{mn}

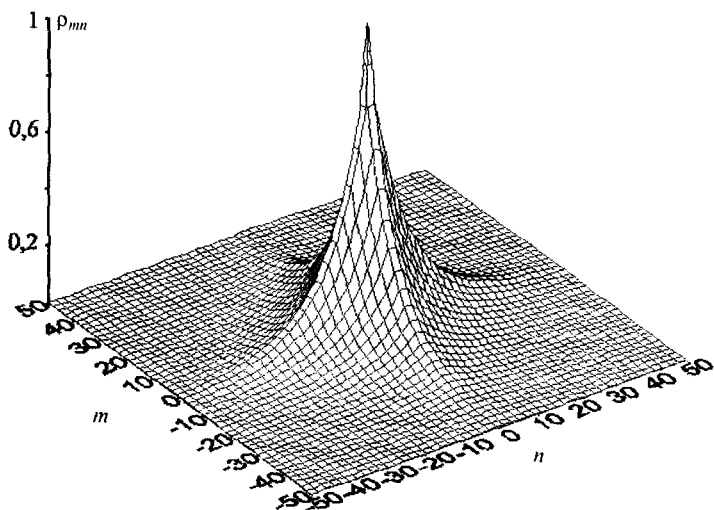


Рис. 2.15. Функция автокорреляции случайного поля (трехточечная модель)

плоскостью, параллельной плоскости $\rho = 0$, имеет вид квадрата или ромба, тогда как для однородного и изотропного поля это круг, а для однородного по строкам и столбцам, но не изотропного — эллипс.

При правильном подборе параметров трехточечная модель позволяет создавать довольно реалистичные изображения, очень похожие на спутниковые изображения земной поверхности (рис. 2.16). Эти изображения можно использовать при исследовании методами математического моделирования на ЭВМ различных алгоритмов обработки изображений.

Параметры, входящие в трехточечную модель, имеют определенный физический смысл: μ_1 — средняя яркость пикселей изображения; σ — среднеквадратическое отклонение, характеризующее контраст. Параметры $1/\alpha$, $1/\beta$ задают характерный размер «пятен» на изображении (в пикселах). При $\rho_1 = \rho_2 = 0,9$ $1/\alpha = 1/\beta = 9,5$ (пикселей).

Гистограмма яркости пикселей изображения на рис. 2.16 имеет вид нормального закона распределения (см. рис. 2.7, а). Приблизиться

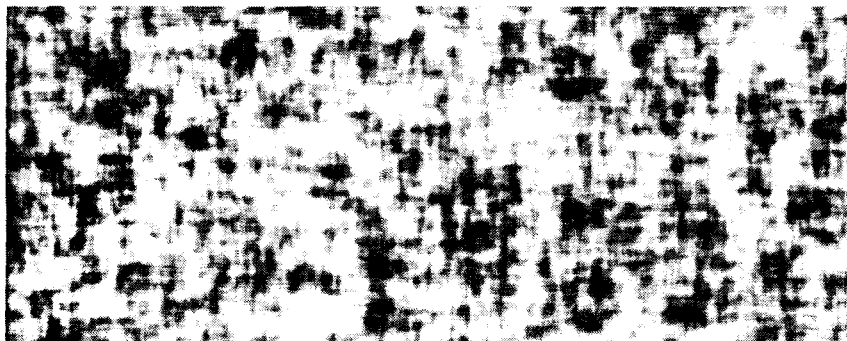


Рис. 2.16. Случайное поле, построенное по трехточечной модели

к однородному (в случае необходимости и изотропному) полю можно, используя более сложные авторегрессионные модели. Четырехточечная модель (см. рис. 2.13, в) записывается как

$$f_{i,j} = \mu_1 + a(f_{i-1,j} - \mu_1) + b(f_{i-1,j-1} - \mu_1) + c(f_{i,j-1} - \mu_1) + d(f_{i+1,j-1} - \mu_1) + eh_{i,j}.$$

При правильном подборе параметров a, b, c, d, e функция R_{mn} значительно отличается от фигуры вращения, т.е. инвариантна относительно поворота вокруг оси, проходящей через точку $i=0, j=0$, а поле практически является однородным и изотропным. Впрочем, изображение, полученное с помощью такой модели, по виду трудно отличить от рис. 2.16.

Схема авторегрессии применяется не только при моделировании случайных полей, но и в других задачах, например при восстановлении пропущенных пикселей в изображении, при подавлении шума на изображении, в некоторых системах кодирования изображений.

2.8.4. Моделирование случайных полей в частотной области

Достоинством схемы авторегрессии являются быстроедействие и малый объем требуемой оперативной памяти. Однако качество изображения не всегда удовлетворительное. Как можно заметить из рис. 2.16, на изображении присутствуют горизонтальные и вертикальные линии. При несколько больших затратах времени есть смысл применить другую модель, обеспечивающую лучшее качество.

Для этого можно использовать случайное поле $f(x, y)$, $a \leq x \leq b$, $c \leq y \leq d$, со статистически независимыми значениями яркости пикселей (см. рис. 2.12), двумерное преобразование Фурье которого

$$F(u, v) = \int_a^b \int_c^d f(x, y) \exp(-iux - ivy) dx dy, \quad (2.14)$$

где u, v — пространственные частоты, $F(u, v)$ — Фурье-образ случайного поля $f(x, y)$. Пусть $K(u, v)$ — коэффициент передачи линейного фильтра, формирующего случайное поле с заданными параметрами. Отметим, что $F(u, v)$ — комплексная, а $K(u, v)$ — действительная величина. Преобразование случайного поля $f(x, y)$ сводится к умножению $F(u, v)$ на $K(u, v)$. Таким образом, Фурье-образ преобразованного поля

$$\Phi(u, v) = F(u, v) K(u, v). \quad (2.15)$$

Моделируемое случайное поле есть обратное преобразование Фурье от $\Phi(u, v)$:

$$S(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(u, v) \exp(iux + ivy) du dv / 2\pi. \quad (2.16)$$

Случайное поле $S(x, y)$ получено путем преобразования исходного поля $f(x, y)$ целиком, в соответствии с рис. 2.10, з, т.е. рассмотренный алгоритм является некаузальным.

Изображение на рис. 2.17, построенное согласно (2.14)–(2.16), где коэффициент передачи $K(u, v) = \exp[-\alpha(u^2 + v^2)]$, напоминает облака. Функция автокорреляции этого поля имеет вид, показанный на рис. 2.9.

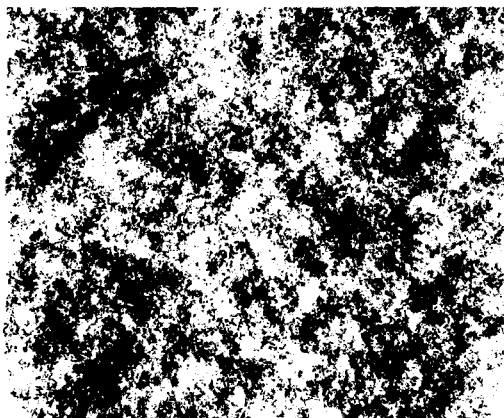


Рис. 2.17. Модельное изображение облаков

2.9. КОНТРОЛЬНЫЕ ЗАДАНИЯ

1. Предложите математические модели линейного и точечного изображений.
2. После идентификации объект представлен в виде черного пятна на белом фоне. Как определить его площадь?
3. Введите изображение (по выбору преподавателя) в ПЭВМ с помощью планшетного сканера.
4. Введите печатный текст в ПЭВМ с помощью планшетного сканера. Используя программу для распознавания текста, оцените достоверность распознавания.
5. С какой скоростью движется электронный луч вдоль строки по экрану размером 30×40 см при скорости кадровой развертки 100 Гц и 800 строках в кадре при воспроизведении линии?
6. Чему равен поперечник луча лазера, обеспечивающего разрешение лазерного принтера в 2000 точек на дюйм?
7. Пусть диагональ монитора составляет 20 дюймов при соотношении высоты и ширины кадра $3/4$ и размере зерна 0,25 мм. Чему равно разрешение такого монитора (число точек по вертикали и горизонтали)?
8. Принтер имеет разрешение 800 точек на дюйм. Каково его разрешение при печати растровым методом изображения с 256 градациями полутонов?
9. Какой объем памяти требуется для хранения одного стандартного кадра размером 185×185 км, полученного с помощью сканера MSS спутника LANDSAT?
10. Используя графический пакет PhotoShop, сожмите изображение (по выбору преподавателя) в формате BMP (групповое кодирование RLE8). Запишите сжатое изображение в память компьютера. Определите размер сжатого изображения. Сравните его с исходным размером.
11. Используя графический пакет PhotoShop, сожмите изображение (по выбору преподавателя) в формате TIF (кодирование LZW) и в формате JPEG, запишите сжатые изображения в память компьютера. Определите размеры изображения в каждом формате. Сравните с размером изображения, сжатого в формате BMP. Каковы ваши выводы?
12. Используя графический пакет PhotoShop, вычитите из исходного изображения такое же изображение, восстановленное из формата JPEG. Чтобы лучше увидеть разность, используйте опции Equalize или Stretch. Проанализируйте полученный результат.
13. Напишите выражение для плотности вероятности закона распределения, равномерного в интервале $[A, B]$.
14. Непрерывная случайная величина f распределена по равномерному закону распределения в интервале $[A, B]$. Чему равны математическое ожидание, дисперсия и медианное значение f ?
15. Составьте программу моделирования случайного поля по двух- и трехточечной авторегрессионной модели.

ЛИТЕРАТУРА

Павлидис Т. Алгоритмы машинной графики и обработки изображений. М.: Радио и связь, 1986. 400 с.

Прэтт У. Цифровая обработка изображений: В 2 книгах. М.: Мир, 1982. 670 с.

Климов А.С. Форматы графических файлов. СПб: НИАП «ДиаСОФТ ЛТД», 1995. 478 с.

Валентюк А.Н., Предко К.Г. Оптическое изображение при дистанционном зондировании. Минск: Навука і тэхніка, 1991. 360 с.

Криксунов Л.З., Падалко Г.А. Тепловизоры. Киев: Техніка, 1987. 166 с.

Шлихт Г.Ю. Цифровая обработка цветных изображений. М.: ЭКОМ, 1997. 336 с.

Ахманов С.А., Дьяков Ю.Е., Чиркин А.С. Введение в статистическую радиофизику и оптику. М.: Наука, 1981. 640 с.

Васильев К.К., Крашенинников В.Р. Методы фильтрации многомерных случайных полей. Саратов: Изд. Саратовского университета, 1990. 126 с.

3. ВОССТАНОВЛЕНИЕ И УЛУЧШЕНИЕ ИЗОБРАЖЕНИЙ

3.1. КАЧЕСТВО ИЗОБРАЖЕНИЙ

Естественная цель обработки изображений — повышение их качества. Но качество изображений — понятие не однозначное. С одной точки зрения, некоторое изображение может считаться высококачественным, с другой — изображением низкогокачественным. Например, изображения, воспроизводимые сканером AVHRR спутника NOAA, существенно уступают по пространственному разрешению изображениям сканера МСУ-Э спутника «Ресурс-О1» (1,1 км против 45 м). Всегда ли это плохо? Существует много задач, где не требуется столь высокого разрешения, но необходима оперативность в получении информации (желательно несколько раз в сутки, а не раз в 1–2 недели) и большое поле обзора. К таким задачам относится, например, мониторинг пожарной опасности в лесах такого обширного региона, каким является Сибирь. В этом случае изображения со спутника NOAA предпочтительнее, чем изображения с «Ресурса». При обработке изображений нередко стремятся устранить все искажения, максимально приблизив преобразованное изображение к идеальному неискаженному. Для этого, например, приходится обрабатывать спутниковое тепловое изображение водоемов или суши, внося коррекцию с учетом свойств датчиков и ослабления ИК-излучения в атмосфере. В результате полет температур оказывается неискаженным, и в этом смысле преобразованное изображение наиболее качественное.

Изменяя контраст теплового изображения (т.е. сознательно искажая его), можно облегчить обнаружение объектов с повышенной температурой. С этой точки зрения отконтрастированное изображение качественнее идеального.

Из этих примеров видно, что повышение качества изображения следует понимать каждый раз в каком-то определенном смысле, с точки зрения конкретной решаемой задачи.

Под *реставрацией изображений* понимают процедуру восстановления искаженного изображения с тем, чтобы приблизить его к идеальному неискаженному. Реставрацию спутниковых изображений земной поверхности реализуют процедуры геометрической и атмосферной коррекции, а также компенсации изменения характеристик фотоприемника. При этом иногда приходится восстанавливать пропущенные данные, заполняя участки, закрытые облаками, пикселями того же вида, что и в окружении этих участков.

Для улучшения изображения его преобразуют в форму, наиболее удобную для визуального или машинного анализа. При этом, как правило, и не пытаются приблизить воспроизводимое изображение к некоторому идеализированному оригиналу. Например, изображение с подчеркнутыми границами (контурами) может субъективно восприниматься лучше, чем неискаженный оригинал, и может оказаться полезным далее при компьютерной обработке. ЭВМ прослеживает контуры объекта, что позволяет определить его форму и размеры. Здесь улучшение изображений используется для того, чтобы подчеркнуть важнейшие признаки изображения и, следовательно, облегчить задачу извлечения информации.

Реставрация и улучшение изображений, а также нанесение координатной сетки и карты гидросети на изображение относят к процедурам *предварительной обработки* космических изображений.

Тематическая обработка космической информации выполняется после предварительной и направлена на решение конкретных задач, на извлечение конкретной информации в той или иной области наук о Земле. Здесь могут быть использованы различные методы и приемы, в зависимости от решаемой задачи. При изучении почв из космоса применяют одни способы, при космическом мониторинге лесных пожаров — другие, при изучении площадей, загрязненных промышленными предприятиями, — третьи. Общим при тематической обработке является использование аппарата распознавания образов.

3.2. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

3.2.1. Геометрическая коррекция спутниковых изображений

В п. 1.4.3 установлено, что для сканеров оптического диапазона с цилиндрической разверткой основной причиной геометрических искажений спутниковых изображений является кривизна поверхности Земли. Специфические искажения, связанные с тем, что наблюдение ведется под углом к надиру, возникают при использовании сканеров с линейной разверткой и радиолокационных станций бокового обзора. Но существуют и другие источники искажений. Солнечно-синхронные орбиты природоведческих спутников не проходят через ось вращения Земли, а наклонены относительно нее. Поэтому, если спутник движется с севера на юг (нисходящий виток орбиты), то вверху изображения будет не север, как на карте, а, например, север-северо-восток. К тому же во время сеанса приема спутниковой информации Земля поворачи-

вается на некоторый угол (за 1 мин на $0,25^\circ$). В принципе такие искажения можно скомпенсировать, если известны проекция орбиты спутника на земную поверхность и механизм искажений. Однако проще воспользоваться другой методикой, эффективной также тогда, когда необходимо обрабатывать архивные изображения, для которых орбитальные данные и угол отклонения оси сканирования от надира неизвестны.

Объекты на спутниковых изображениях бывает необходимо сопоставлять с географической картой (осуществить географическую привязку спутниковых данных) для определения географических координат объектов. Географическую привязку и геометрическую коррекцию можно объединить в одну операцию совмещения деталей спутникового изображения и карты. Пусть система координат (x, y) соответствует спутниковому изображению, а система (u, v) — карте. Требуется найти преобразование $u_k = f(x_k, y_k)$, $v_k = g(x_k, y_k)$, устанавливающее соответствие между положением k -го пиксела на изображении и географическими координатами. Так как вид функций f и g заранее неизвестен, то применяется полиномиальная аппроксимация. Обычно используют полиномы второй степени:

$$\begin{aligned} u_k &= a_0 + a_1 x_k + a_2 y_k + a_3 x_k y_k + a_4 x_k^2 + a_5 y_k^2; \\ v_k &= b_0 + b_1 x_k + b_2 y_k + b_3 x_k y_k + b_4 x_k^2 + b_5 y_k^2. \end{aligned} \quad (3.1)$$

Первые члены с коэффициентами a_0, b_0 ответственны за сдвиг изображения по x и по y , члены с коэффициентами a_1, a_2, b_1, b_2 — за линейное изменение масштаба по x и по y , члены с a_3 и b_3 — за вращение изображения, члены с a_4, a_5, b_4, b_5 — за нелинейное изменение масштаба.

Коэффициенты a_i и b_i определяются из решения системы (3.1). На изображении и на карте отыскивают одинаковые точки (называемые контрольными — control points, reference points), их координаты подставляют в уравнения. В качестве контрольных точек удобно использовать элементы гидросети: устья рек, мысы, крутые изгибы русла рек и т.п. Количество точек должно быть достаточным для решения уравнений. Для полиномов второго порядка можно ограничиться шестью контрольными точками, но желательно, чтобы их число достигало 15–20 с распределением по всему полю, это позволяет использовать метод наименьших квадратов и сделать оценку коэффициентов менее зависимой от ошибок в определении координат на изображении и на карте.

Примером может служить осеннее радиолокационное изображение района к востоку от Норильска (рис. 3.1, а, спутник ERS-2), на котором



a)



б)

Рис. 3.1. Радиолокационное изображение местности к востоку от Норильска (а) и фрагмент топографической карты этого района (б)

отмечены контрольные точки и карта района (рис. 3.1, б), по которой определены координаты опорных точек:

1 — $89^{\circ},5726$ в. д. $69^{\circ},4241$ с. ш.; 2 — $89^{\circ},7894$ в. д. $69^{\circ},3506$ с. ш.;
3 — $89^{\circ},6436$ в. д. $69^{\circ},2797$ с. ш.; 4 — $89^{\circ},8934$ в. д. $69^{\circ},2337$ с. ш.;
5 — $90^{\circ},5322$ в. д. $69^{\circ},2106$ с. ш.; 6 — $90^{\circ},8033$ в. д. $69^{\circ},4069$ с. ш.;
7 — $90^{\circ},9876$ в. д. $68^{\circ},9968$ с. ш.; 8 — $91^{\circ},4512$ в. д. $69^{\circ},0269$ с. ш.;
9 — $89^{\circ},6608$ в. д. $68^{\circ},8202$ с. ш.; 10 — $90^{\circ},1147$ в. д. $68^{\circ},8138$ с. ш.;
11 — $90^{\circ},0117$ в. д. $68^{\circ},6736$ с. ш.; 12 — $90^{\circ},6759$ в. д. $68^{\circ},6895$ с. ш.;
13 — $91^{\circ},3102$ в. д. $68^{\circ},7966$ с. ш.

За опорные точки принимаются мысы и водные заливы. Координаты опорных точек на изображении во внутренней системе координат (первая цифра — координата по x , вторая — по y):

1 — 109 и 23; 2 — 149 и 56; 3 — 124 и 89; 4 — 161 и 117; 5 — 291 и 128; 6 — 344 и 25; 7 — 376 и 241; 8 — 461 и 222; 9 — 116 и 329; 10 — 214 и 374; 11 — 296 и 404; 12 — 214 и 395; 13 — 431 и 340.

Решение системы (3.1) дает:

$a_0 = 88,9523$; $b_0 = 69,4845$; $a_1 = 5,6923 \cdot 10^{-3}$; $b_1 = -1,6122 \cdot 10^{-4}$;
 $a_2 = 5,1447 \cdot 10^{-4}$; $b_2 = -1,98888 \cdot 10^{-3}$; $a_3 = 3,402 \cdot 10^{-7}$; $b_3 = 8,47 \cdot 10^{-8}$;
 $a_4 = -1,033 \cdot 10^{-6}$; $b_4 = 2,268 \cdot 10^{-7}$; $a_5 = -1,1102 \cdot 10^{-6}$; $b_5 = 2,829 \cdot 10^{-8}$.

Средний квадрат ошибки в определении новых координат $\epsilon = 0,0135$. Следует так выбирать число контрольных точек и их расположение, чтобы достичь минимума ϵ .

Результат геометрической коррекции и топографической привязки с использованием (3.1) можно видеть на рис. 3.2.

3.2.2. Радиометрическая коррекция результатов дистанционного зондирования

Измерительная аппаратура природоведческих спутников Земли перед запуском тщательно проверяется и калибруется, кроме того, информация со спутника в течение некоторого времени после его запуска (до нескольких месяцев) проходит верификацию. В результате данные дистанционного зондирования могут быть надежно использованы для решения различных практических задач.

За время функционирования спутников на орбите измерительная аппаратура деградирует под воздействием неблагоприятных факторов космического пространства. Поэтому показания датчиков сканеров

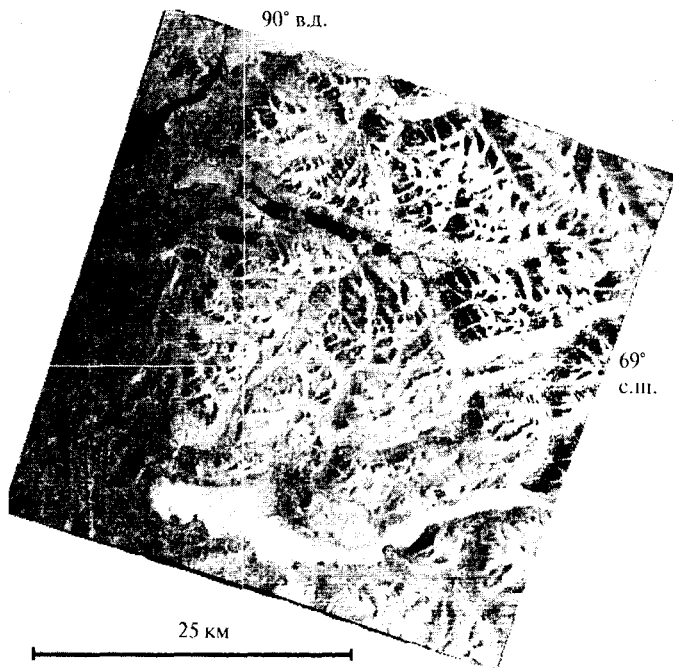


Рис. 3.2. Радиолокационное изображение после геометрической коррекции

необходимо подвергать так называемой *радиометрической коррекции*. Рассмотрим ее на примере обработки данных сканера AVHRR спутника NOAA.

Сканер AVHRR имеет по одному каналу видимого и ближнего инфракрасного диапазона (1-й и 2-й каналы), один ИК-канал на 3,5 мкм (3-й) и два канала теплового ИК-диапазона (4-й и 5-й). Предусмотрена бортовая калибровка последних трех каналов. Для этого сканер направляется на космическое пространство (эта точка принимается за нуль) и в полость абсолютно черного тела, установленного на борту, что дает две точки для коррекции температуры по линейному закону. Корректировочные коэффициенты для 3–5-го каналов включены в файл данных, передаваемых со спутника. 1-й и 2-й каналы калибруются только на Земле перед запуском спутника (соответствующие коэффициенты также включены в файл данных). С течением времени показания 1-го и 2-го каналов необходимо корректировать. Для этого наблюдаются одни и те же объекты на Земле, текущая интенсивность излучения сравнивается с результатами предыдущих наблюдений. И текущие, и полученные прежде данные подвержены разбросу из-за влия-

яния атмосферы, прозрачность которой постоянно меняется из-за естественной изменчивости природных объектов и других факторов. Корректировочные коэффициенты для 1-го и 2-го каналов определяются путем статистической обработки и ежемесячно обновляются, их можно найти в сети Интернет: <http://140.90.207.25:8080/EBV/ml/nical.html>.

В качестве примера рассмотрим процедуру калибровки и корректировки данных 4-го канала. Показания U датчика этого канала после аналого-цифрового преобразования лежат в пределах 0–1023 (10-битное кодирование). Пусть $U = 427$. Поток энергии можно найти по формуле $B(\lambda, T) \cdot \lambda^2 = A + K \cdot U$, $A = 590,888 \text{ Вт} \cdot \text{мкм}/\text{м}^2$, $K = -1,60156 \text{ Вт} \cdot \text{мкм}/\text{м}^2$ — корректировочные коэффициенты из файла данных. В нашем примере $B(\lambda, T) = 907,022/\lambda^2 \text{ Вт}/\text{м}^2\text{мкм}$. Из формулы Планка (1.1) получаем, что при $\lambda = 10,96 \text{ мкм}$, $c_1 = 1,1911 \cdot 10^8 \text{ Вт} \cdot \text{мкм}^4/\text{м}^2$, $c_2 = 14\,388 \text{ мкм} \cdot \text{К}$ радиационная температура объекта

$$T = (\lambda/c_2) \ln(c_1/\lambda^5 B + 1) = 284,57 \text{ К}.$$

Пусть $U = 428$, т.е. показания датчика изменились на единицу. В этом случае $B = 905,421/\lambda^2 \text{ Вт}/\text{м}^2\text{мкм}$, $T = 284,46 \text{ К}$. Таким образом, изменению показаний датчика $\Delta U = 1$ соответствует изменение температуры $\Delta T = 0,1 \text{ К}$. Величина ΔT характеризует *радиационное разрешение* датчика.

3.2.3. Атмосферная коррекция

Задача атмосферной коррекции — одна из самых сложных из задач реставрации результатов дистанционного зондирования Земли. В первую очередь это связано с тем, что, как правило, отсутствует информация об оптической толщине τ атмосферы над интересующими объектами. Обычно космические изображения суши, значительную часть которых занимает облачность, выбраковываются. Нередко дальнейшая обработка оставшихся изображений ведется без атмосферной коррекции.

Лучшим выходом из положения была бы установка по всей поверхности суши обширной сети солнечных спектрофотометров, измеряющих интенсивность солнечного излучения I , прошедшего через атмосферу, в различных участках спектра. Зная интенсивность I_0 за пределами атмосферы, по закону Бугера $I = I_0 \exp(-\tau \sec \delta)$ можно оценить τ для разных длин волн оптического диапазона: $\tau = (1/\sec \delta) \ln I_0/I$. Здесь δ — зенитный угол Солнца (см. рис. 1.7).

Поскольку такой сети не существует, приходится довольствоваться данными немногих спектрофотометров или прибегать к косвенным методам коррекции.

Иначе дело обстоит с атмосферой над морями и океанами. В красном и ИК-участке спектра поверхность воды по своим оптическим характеристикам близка к абсолютно черному телу. Существенно больший коэффициент отражения и рассеяния имеют туманы, дымки, облака, их хорошо видно на фоне воды. Это позволяет оценить оптическую толщину τ . Данные об оптической толщине над морями и океанами помещены в сети Интернет по адресу: http://las.saa.noaa.gov/las-bin/climate_server/, их можно использовать для коррекции спутниковых изображений прибрежных районов.

Изображение части Хабаровского края (слева), Татарского пролива и о. Сахалин (справа), полученное 14 августа 1998 г. (1-й канал сканера AVHRR спутника NOAA) (рис. 3.3, а), очень низкого качества. В это время на территории Хабаровского края и на Сахалине бушевали обширные лесные пожары, весь район был сильно задымлен. После атмосферной коррекции с использованием закона Бугера (значения $\tau = 0,03-0,23$ взяты из сети Интернет, угол δ изменялся от -8° до 14°) (рис. 3.3, б) просматривается Татарский пролив и несколько лучше, чем на рис. 3.3, а, — о. Сахалин. Для Хабаровского края использованы значения τ для Татарского пролива и Охотского моря, при этом результат, как и следовало ожидать, оказался неудовлетворительным: в левом верхнем углу появилась темная область, похожая на поверхность моря, хотя на самом деле это суша.

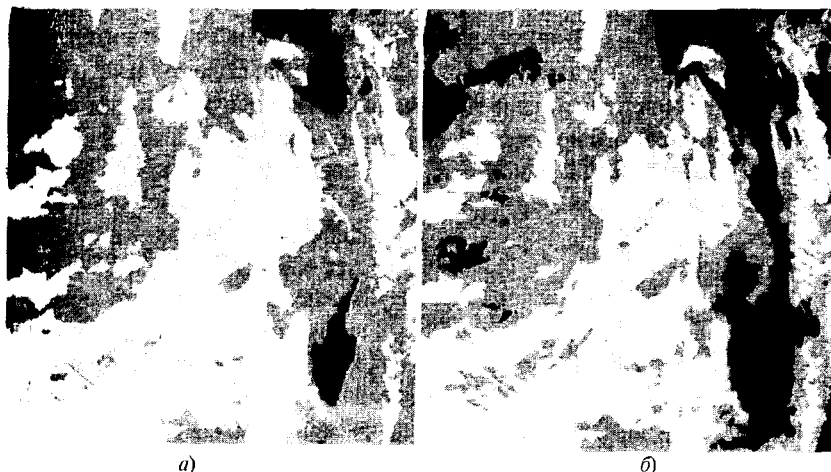


Рис. 3.3. Атмосферная коррекция изображения прибрежных районов

Существует достаточно много оценочных методов учета загрязнения атмосферы и косвенных методов атмосферной коррекции. Можно, например, оценить поглощение излучения молекулами воды по эквивалентной массе водяного пара в атмосфере, вычисленной по результатам измерений температуры и влажности. Можно также использовать усредненные сезонные значения τ для данной местности.

Алгоритм порогового выделения загрязненных областей оказывается наиболее простым, но достаточно эффективным. Его идея основана на выделении границы между «загрязненной» и «чистой» областями, после чего все значения изображения в пределах этой границы исключаются из рассмотрения. Порог для 1-го и 2-го каналов сканера AVHRR спутника NOAA можно выбрать по нормализованному вегетационному индексу $NDVI$ для всего изображения, который вычисляется по чувствительности $NDVI$ к присутствию облаков и аэрозолей на изображении. Для растений в нормальном состоянии $NDVI$ близок к 0,6–0,65, для растений в угнетенном состоянии — к 0,3–0,4. В 1-м и 2-м каналах облака и аэрозоли имеют приблизительно одинаковую яркость, поэтому $NDVI$ близок к 0. Таким образом, пороговыми являются низкие значения $NDVI$. Порог для 3–5-го каналов выбирается как среднее значение температуры поверхности и облаков.

Кроме солнечного излучения, отраженного и рассеянного поверхностью, сканеры спутника принимают солнечное излучение, рассеянное на молекулах газов и на аэрозолях. Интенсивность молекулярного рассеяния обратно пропорциональна длине волны в четвертой степени, эффект наиболее заметен в коротковолновой части спектра, который ответствен за голубой цвет неба. Рассеяние на аэрозолях (размер частиц от $0,1\lambda$ до 10λ) приводит к зависимости интенсивности от длины волны $\lambda^{-\alpha}$, $0 < \alpha < 4$. Размеры частиц дыма, пыли, облаков много больше длины волны видимого и ИК-диапазонов, где интенсивность рассеянного солнечного излучения не зависит от длины волны. Рассеяние в атмосфере приводит к дополнительной засветке, заметной в 1-м и 2-м каналах сканера AVHRR спутника NOAA и еще более заметной в каналах

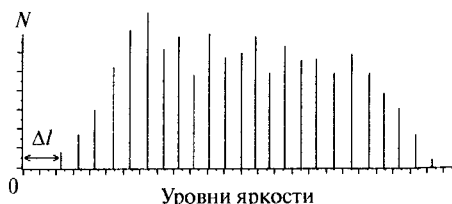


Рис. 3.4. Гистограмма уровней яркости при наличии рассеяния в атмосфере

лах сканеров спутников «Ресурс-О», LANDSAT, SPOT, работающих в зеленой области видимого диапазона. Засветка, различная для разных каналов, способна привести, например, к ошибкам в оценке вегетационного индекса растительного покрова.

Влияние засветок за счет рассеяния в атмосфере можно частично устранить, если использовать следующую процедуру. Как правило, на всех изображениях земной поверхности присутствуют пиксели с нулевой или близкой к нулю яркостью. Рассеяние в атмосфере приводит к увеличению яркости, гистограмма смещается вправо (рис. 3.4). Методика коррекции состоит в том, что из всех уровней яркости вычитается величина ΔI , определяемая по гистограмме.

3.2.4. Восстановление пропущенных пикселов

Часто встречающимися на практике помехами при исследовании поверхности Земли из космоса являются облака. Если их слишком много, изображение не годится для анализа; если площадь, покрытая облаками, невелика, а облака небольшие, то области под облаками на изображении можно восстановить путем интерполяции (экстраполяции) с использованием уравнения авторегрессии. Конечно, таким образом невозможно получить, например, изображение невидимого населенного пункта, но заполнить пустое место в изображении лесного массива пикселями такой же структуры, что и окружение, вполне возможно. В системах обработки космической информации на изображения накладываются координатная сетка и кресты для определения координат (их можно заметить на рис. 3.5). При некоторых видах тематической обработки, когда проводится классификация изображений, эти линии, как и облака, являются помехами.

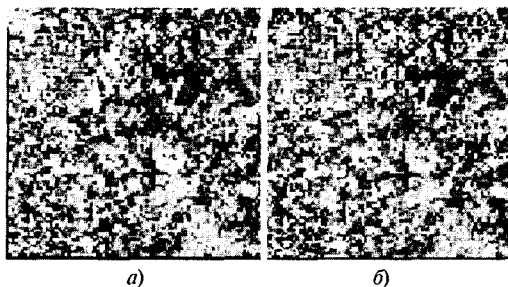


Рис. 3.5. Исходное изображение (а) и скорректированное восстановлением пропущенных пикселов (б)

Запишем уравнение авторегрессии в виде

$$g_{i,j} = a_0 f_{k,m} + a_1 f_{k+1,m} + a_2 f_{k+1,m+1} + \dots + a_{s+m} f_{k+s,m+q} + b \xi_{ij}.$$

Здесь $g_{i,j}$ — прогнозируемая яркость пропущенного пиксела (оценка яркости пиксела), $a_{i,j}$, b — некоторые коэффициенты; важно, чтобы известные яркости $f_{h,q}$ пикселей из окружения пропущенного пиксела не повторялись.

Потребуем, чтобы среднее значение $\langle g_{i,j} \rangle$ совпадало с истинным средним значением яркости $\langle f_{i,j} \rangle$ (условие несмещенности оценки). Для однородного и изотропного (хотя бы в пределах $(s+q)$ — окрестности пропущенного пиксела) поля

$$\langle f_{i,j} \rangle = \langle f_{r,p} \rangle = m_1,$$

где m_1 — средняя яркость любого пиксела из этой окрестности, эта величина одинакова для всех пикселей окрестности. Учитывая, что $\langle \xi_{ij} \rangle = 0$, получаем $\sum_{r,p} a_{r+p} m_1 = m_1$, откуда находим уравнение нормировки:

$$\sum_{r,p} a_{r+p} = 1. \quad (3.2)$$

Обозначим $g_{i,j} = g_{i,j} - m_1$, $f_{k,m} = f_{k,m} - m_1$. Вычитая (3.2) из уравнения авторегрессии, получаем

$$g_{i,j} = a_0 f_{k,m} + a_1 f_{k+1,m} + a_2 f_{k+1,m+1} + \dots + a_s f_{k+s,m+q} + b \xi_{ij}. \quad (3.3)$$

Остальные уравнения относительно a_q можно получить из условия минимума D — среднего квадрата погрешности восстановления яркости пропущенного пиксела, в этом случае $g_{i,j}$ — несмещенная и эффективная оценка, а уравнение авторегрессии обеспечивает оптимальную, в смысле минимума среднего квадрата ошибки, линейную процедуру оценивания.

Следует ожидать, что оценка $g_{i,j}$ будет отличаться от истинного (но неизвестного) значения $f_{i,j}$ на величину $\varepsilon_{i,j}$, которую можно назвать погрешностью оценки:

$$\varepsilon_{i,j} = g_{i,j} - f_{i,j} = g_{i,j} - f_{i,j}.$$

Средний квадрат погрешности

$$D = \langle [g_{i,j} - f_{i,j}]^2 \rangle = \langle [\sum a_{r,p} f_{r,p} - f_{i,j}]^2 \rangle = \sum \sum a_{r+p} a_{r+p} \langle f_{r,p} f_{m,n} \rangle - 2 \sum a_{r+p} \langle f_{r,p} g_{m,n} \rangle + \langle g_{i,j}^2 \rangle + b^2 \langle \xi_{i,j}^2 \rangle - 2b \langle \xi_{i,j} g_{i,j} \rangle. \quad (3.4)$$

Здесь учтено, что $\langle \xi_{i,j} f_{r,p} \rangle = 0$. Средний квадрат погрешности минимален при $\partial D / \partial a_{r+p} = 0$, $r, p = 1, 2, \dots, N$, $\partial D / \partial b = 0$. Таким образом, $b = 0$, т.е. в данной задаче член с $\xi_{i,j}$ исключается. Величины a_0, a_1, a_2, \dots есть

решение системы линейных алгебраических уравнений, коэффициентами которых служат коэффициенты корреляции $\langle f_{r,p} f_{m,n} \rangle$.

На практике при решении системы используются оценки коэффициентов корреляции, вычисляемые по $(s + q)$ известным яркостям пикселей из окружения пропущенного пиксела. Далее a_0, a_1, a_2, \dots пересчитываются так, чтобы они удовлетворяли уравнению (3.2).

Рассмотренная процедура, обычно называемая процедурой Криге (Krige) или кригингом, может применяться при обработке случайных полей, когда требуется перейти к регулярной сетке, хотя значения поля заданы на сетке со случайно расположенными узлами. Процедура позволяет также перейти от сетки одного формата к сетке другого формата (см. п. 5.9.5).

3.2.5. Улучшение изображений путем изменения контраста

Слабый контраст — наиболее распространенный дефект фотографических, сканерных и телевизионных изображений, обусловленный ограниченностью диапазона воспроизводимых яркостей. Под контрастом обычно понимают разность максимального и минимального значений яркости. Путем цифровой обработки контраст можно повысить, изменяя яркость каждого элемента изображения и увеличивая диапазон яркостей. Для этого разработано несколько методов.

Пусть, например, уровни некоторого черно-белого изображения занимают интервал 6—158 со средним значением яркости 67 при возможном наибольшем интервале значений 0—255. Гистограмма яркостей исходного изображения на рис. 3.6, а показывает, сколько пикселей N

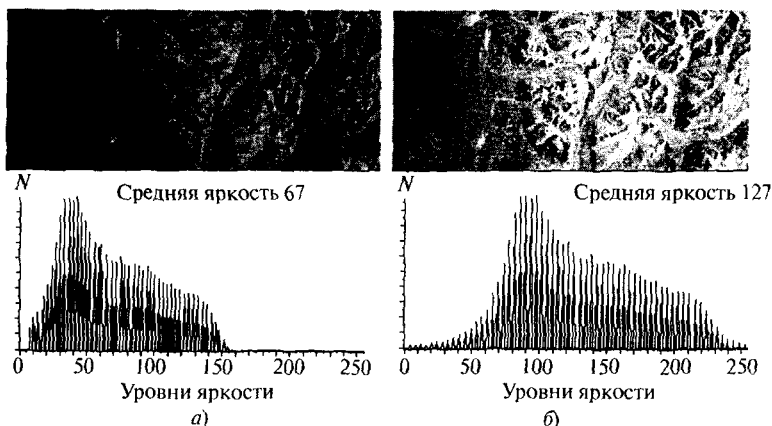


Рис. 3.6. Исходное изображение (а) и изображение после линейной растяжки гистограммы (б)

с близким значением яркости f попадает в интервал от f_i до $f + \Delta f_i$. На этом малококонтрастном изображении превалирует темный оттенок. Возможным методом улучшения контраста может стать так называемая *линейная растяжка гистограммы* (stretch), когда уровням исходного изображения, лежащим в интервале $[f_{\min}, f_{\max}]$, присваиваются новые значения, с тем чтобы охватить весь возможный интервал изменения яркости, в данном случае $[0, 255]$. При этом контраст существенно увеличивается (рис. 3.6, б). Преобразование уровней яркости осуществляется по формуле

$$g_i = c + df_i, \quad (3.5)$$

где f_i — старое значение яркости i -го пиксела, g_i — новое значение, c, d — коэффициенты. Для рис. 3.6, а $f_{\min} = 6, f_{\max} = 158$. Выберем c и d таким образом, чтобы $g_{\min} = 0, g_{\max} = 255$. Из (3.5) получаем $c = -10,01; d = 1,67$.

Контраст можно еще более улучшить, используя *нормализацию гистограммы*, когда на весь максимальный интервал уровней яркости $[0, 255]$ растягивается не вся гистограмма от f_{\min} до f_{\max} , а ее наиболее интенсивный участок (f'_{\min}, f'_{\max}), исключая из рассмотрения малоинформативные «хвосты». На рис. 3.7, б исключено 5% пикселей.

Целью выравнивания гистограммы (эту процедуру называют также *линеаризацией* и *эквализацией* — equalization) является такое преобразование, чтобы, в идеале, все уровни яркости приобрели одинаковую ча-

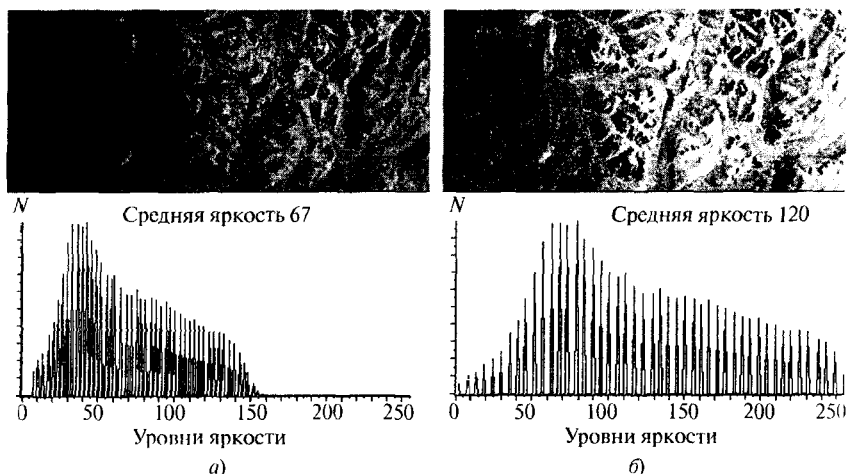


Рис. 3.7. Нормализация гистограммы

стоту, а гистограмма яркостей отвечала равномерному закону распределения (рис. 3.8).

Пусть при числе уровней квантования яркости J изображение имеет формат из N пикселей по горизонтали и M — по вертикали. Общее число пикселей равно $N \cdot M$, на один уровень яркости попадает в среднем $n_0 = N \cdot M/J$

пикселей. Например, $N = M = 512$, $J = 256$. В этом случае $n_0 = 1024$.

Расстояние Δf между дискретными уровнями яркости от f_i до f_{i+1} в гистограмме исходного изображения одинаковое, но на каждый уровень выпадает разное число пикселей. При эквализации гистограммы расстояние Δg_i между уровнями g_i и g_{i+1} различно, но число пикселей на каждом уровне в среднем одинаковое и равно n_0 . Алгоритм эквализации несложен. Пусть уровнями с малой яркостью обладает небольшое количество пикселей, как на рис. 3.9, а. Например, уровень яркости 0 на исходном изображении имеют 188 пикселей, уровень 1 — 347 пикселей, уровень 2 — 544 пикселя. В сумме это 1079 пикселей, т.е. приблизительно n_0 . Присвоим всем этим пикселям уровень 0. Пусть на исходном изображении число пикселей с уровнями яркости 3 и 4 в сумме приблизительно также равно n_0 . Этим пикселям присваивается уровень 1.

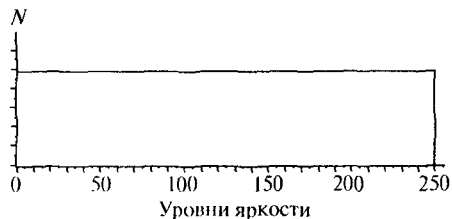


Рис. 3.8. Гистограмма, отвечающая равномерному закону распределения

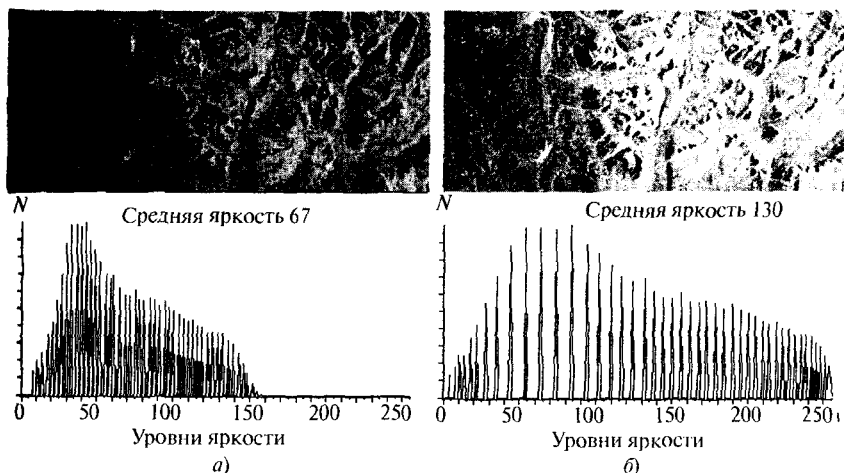


Рис. 3.9. Эквализация гистограммы

С другой стороны, пусть число пикселей с уровнем 45 на исходном изображении составляет 3012, т.е. приблизительно $3n_0$. Всем этим пикселям присваивается некоторый одинаковый уровень g_j , не обязательно равный 45, а соседние два уровня остаются *незаполненными*. Эти процедуры повторяются для всех уровней яркости. Результат эквализации можно видеть на рис. 3.9, б.

В каждом конкретном случае выбирают ту процедуру преобразования гистограмм, которая приводит к наилучшему, с точки зрения пользователя, результату.

Процедуру видоизменения гистограммы можно рассматривать как попиксельное преобразование входной яркости f_j , $f_0 = f_j = f_j$, в выходную яркость g_k , $g_0 = g_k = g_k$, в результате которого исходное распределение вероятностей $P\{f_j\}$ переходит в распределение вероятностей $P\{g_k\}$, имеющее желаемую форму. Очевидно, что сумма вероятностей яркостей всех пикселей должна равняться единице:

$$\sum_{j=0}^J P\{f_j\} = 1, \quad \sum_{k=0}^K P\{g_k\} = 1.$$

Вероятность попадания исходной яркости f_j в интервал $0-m$ должна равняться вероятности попадания яркости преобразованного изображения g_k в интервал $0-n$ для всех $m \leq J$, $n \leq K$:

$$\sum_{j=0}^m P\{f_j\} = \sum_{k=0}^n P\{g_k\}.$$

В случае конкретного изображения распределение в левой части заменяют на гистограмму $H(f_j)$, поэтому

$$\sum_{j=0}^m H\{f_j\} = \sum_{k=0}^n H\{g_k\}.$$

Решая это уравнение, можно найти требуемое преобразование $g_k = T\{f_j\}$. Решение записывается в виде таблицы, в которой для каждого входного уровня f_j указывается соответствующий выходной уровень g_k .

3.3. ЛИНЕЙНАЯ ПРОСТРАНСТВЕННО-ИНВАРИАНТНАЯ ФИЛЬТРАЦИЯ

Реальные изображения наряду с полезной информацией содержат различные помехи. Источниками помех являются собственные шумы фотоприемных устройств, зернистость фотоматериалов, шумы каналов связи. Наконец, возможны геометрические и радиометрические искажения, изображение может быть расфокусировано (но расфокусировка не типична для спутниковых изображений с разрешением 10 м и более); для изображений с разрешением 1 м и менее турбулентность атмосферы приводит к размыванию мелких деталей при коротких экспозициях; при экспозициях в несколько секунд искажения можно описать первым членом ряда (3.6) при $h_1(x, y) \sim \exp[-(x^2 + y^2)/\alpha]$.

Модель искаженного помехами непрерывного изображения имеет вид

$$f(x, y) = m(x, y) \cdot Fs(x, y) + n(x, y),$$

где $f(x, y)$ — искаженное изображение, $m(x, y)$ — мультипликативная помеха, модулирующая изображение по яркости, $s(x, y)$ — исходное изображение, F — функционал, описывающий геометрические и радиометрические искажения, а также расфокусировку, $n(x, y)$ — аддитивная помеха, накладывающаяся на изображение.

Модуляция спутникового изображения по яркости может происходить из-за того, что атмосфера над различными точками Земли имеет различную прозрачность, восходящее излучение от этих точек проходит различный путь в атмосфере.

При реставрации изображений необходимо восстановить исходное изображение. Уже рассмотрены методы устранения геометрических, радиометрических искажений, атмосферной коррекции, восстановления пропущенных пикселей. Будем считать, что эти искажения отсутствуют, $m(x, y) = 1$. Таким образом, $f(x, y) = Fs(x, y) + n(x, y)$.

Результат реставрации $s(x, y) = g(x, y)$ запишем как следствие воздействия на $f(x, y)$ некоторого оператора: $g(x, y) = Tf(x, y)$, где системный оператор T указывает на правило, по которому «входному сигналу» $f(x, y)$ ставится в соответствие «выходной сигнал» $g(x, y)$. Для того чтобы модель была полной, необходимо также указать области допустимых значений $f(x, y)$ и $g(x, y)$. При реставрации применяют оператор T , минимизирующий расстояние между $g(x, y)$ и $s(x, y)$ при заданных статистических характеристиках случайных полей $s(x, y)$, $n(x, y)$ и известном F . В качестве критерия близости $g(x, y)$ и $s(x, y)$ часто используют критерий минимума среднеквадратической ошибки: $\min \langle [g(x, y) - s(x, y)]^2 \rangle$.

В задачах улучшения изображений обычно считается, что $n(x, y) = 0$, функцией оператора T является сглаживание резких перепадов яркости, подчеркивание или выделение контуров и т.п.

Будем рассматривать пространственно-инвариантные операторы, выходная реакция которых не зависит от изменения начала отсчета по x и по y и от ориентации объектов на изображении. Первое условие означает, что оператор переводит однородное случайное поле в однородное. Второе условие означает, что оператор переводит изотропное поле в изотропное. Отметим, что свойства пространственной инвариантности выполняются строго, если области допустимых значений координат x, y попадают в интервал $(-\infty, \infty)$. Реальные изображения имеют конечные размеры, $A \leq x \leq B$; $C \leq y \leq D$, условие пространственной инвариантности выполняется приближенно.

Оператор называется линейным, если для него справедлив принцип суперпозиции — реакция на сумму сигналов $f_1(x, y)$ и $f_2(x, y)$ равна сумме реакций на каждое из воздействий в отдельности, т.е.

$$T(f_1(x, y) + f_2(x, y)) = Tf_1(x, y) + Tf_2(x, y).$$

Для любого произвольного числа α справедливо

$$T\alpha f(x, y) = \alpha Tf(x, y).$$

Свойства линейности выполняются строго, если области допустимых значений яркости f, g попадают в интервал $(-\infty, \infty)$. При цифровой обработке яркость — величина вещественная, неотрицательная и ограниченная, обычно $0 \leq f, g \leq 255$. Если каждому $g(x, y)$ отвечает единственное $f(x, y)$, то оператор T может быть представлен в виде функционального степенного ряда (ряда Вольтерра):

$$g(x, y) = \iint f(x', y') h_1(x, y, x', y') dx' dy' + \iint \iint f(x'_1, y'_1) \times f(x'_2, y'_2) h_2(x_1, y_1, x'_1, y'_1, x_2, y_2, x'_2, y'_2) dx'_1 dy'_1 dx'_2 dy'_2. \quad (3.6)$$

Здесь интегрирование ведется по всей области, где определены x, y ; записаны два члена ряда Вольтерра (линейный и квадратичный); весовые множители $h_1(x, y, x', y')$ и $h_2(x_1, y_1, x'_1, y'_1, x_2, y_2, x'_2, y'_2)$ называются ядрами Вольтерра первого и второго порядка.

Выражение (3.6), где интегрирование ведется по всей области определения x и y , характеризует преобразование всего изображения целиком — глобальную фильтрацию. Можно обрабатывать изображение по частям, в этом случае осуществляется локальная фильтрация.

Ядро первого порядка $h_1(x, y, x', y')$ в оптике именуют функцией рассеяния точки (ФРТ) (рис. 3.10). Такое изображение точечного источника на выходе оптической системы уже является не точкой, а некоторым пятном. В соответствии с (3.6) все точки изображения $f(x', y')$ превращаются в пятна, которые и суммируются (интегрируются). Не следует думать, что эта процедура обязательно приводит к расфокусировке изображения, наоборот, можно подобрать такую ФРТ, которая позволит сфокусировать расфокусированное изображение.

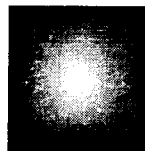


Рис. 3.10. Пример функции рассеяния точки

Чтобы для ФРТ выполнялось условие пространственной инвариантности, т.е. чтобы ФРТ не изменялась при изменении начала отсчета по x и по y , она должна иметь вид $h_1(x, y, x', y') = h_1(x - x', y - y')$. В этом случае $h_1(x, y, x', y') = h_1(x + x_0, y + y_0, x' + x_0, y' + y_0)$. Кроме того, ФРТ должна обладать осевой симметрией.

При обработке растровых изображений на прямоугольной сетке проще всего реализовать ФРТ конечных размеров в виде прямоугольной матрицы форматом $N \times N$, например, 3×3 :

$$\begin{vmatrix} a_{11} & a_{12} & a_{11} \\ a_{12} & a_{22} & a_{12} \\ a_{11} & a_{12} & a_{11} \end{vmatrix}, \quad (3.7)$$

только три элемента которой независимы, в этом случае матрица инвариантна относительно поворотов, кратных 90° . Опыт обработки изображений показывает, что отсутствие более строгой осевой симметрии ФРТ слабо сказывается на результатах. Иногда используют 8-угольные матрицы, инвариантные относительно поворотов на 45° .

3.3.1. Линейные преобразования в частотной плоскости

Важнейшей особенностью линейного оператора является то обстоятельство, что он изменяет не форму входного синусоидального сигнала $s(t) = A \cos(\omega t + \varphi)$, а только амплитуду A и фазу φ , хотя форма несинусоидального сигнала может сильно измениться. С математической точки зрения синус и косинус являются собственными функциями линейной системы. Это обусловило широкое использование интеграла Фурье и ряда Фурье при линейной обработке сигналов и при обработке изображений. В последнем случае существенно упрощается процедура глобальной фильтрации.

Пусть $f(x, y)$ — функция двух переменных, определенная на интервалах $(-\infty < x < \infty)$, $(-\infty < y < \infty)$ и удовлетворяющая условию абсолютной интегрируемости

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x, y)| dx dy < \infty.$$

Тогда существует интеграл Фурье, это означает следующее:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp(-2\pi i[ux + vy]) dx dy, \quad (3.8)$$

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) \exp(-2\pi i[ux + vy]) du dv, \quad (3.9)$$

Комплексная экспонента является линейной комбинацией синуса и косинуса: $\exp(i 2\pi[ux + vy]) = \cos(2\pi[ux + vy]) + i \sin(2\pi[ux + vy])$, где i — мнимая единица.

Выражение (3.8) носит название прямого преобразования Фурье, а (3.9) — обратного. Преобразование Фурье линейное, так как интеграл — линейная функция. Переменные x и y — это координаты; u и v называются пространственными частотами, а функция $F(u, v)$ — спектром пространственных частот или спектром. Используя преобразование Фурье, переходим от координатной плоскости (x, y) к частотной (u, v) . Такой переход имеет смысл, так как некоторые свойства спектра проще, чем функции $f(x, y)$, описывающей распределение яркости в координатной плоскости. Пусть, например, требуется найти результат глобального линейного преобразования некоторого изображения:

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') h(x - x', y - y') dx' dy'. \quad (3.10)$$

В координатной плоскости для этого требуется вычислить интеграл типа свертки (3.10), что часто достаточно сложно. Если ввести частотный коэффициент передачи $K(u, v)$, который связан с ФРТ парой преобразований Фурье:

$$K(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(x, y) \exp(-2\pi i[ux + vy]) dx dy, \quad (3.11)$$

$$h_1(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(u, v) \exp(2\pi i\{ux + vy\}) du dv, \quad (3.11')$$

то в плоскости пространственных частот (3.10) сведется к перемножению функций $F(u, v)$ и $K(u, v)$:

$$G(u, v) = F(u, v) \cdot K(u, v), \quad (3.12)$$

где $G(u, v)$ — спектр после линейного преобразования.

Спектр $F(u, v)$ от вещественной функции $f(x, y)$, вообще говоря, является комплексной функцией. Для пространственно-инвариантных ФРТ частотный коэффициент передачи $K(u, v)$ всегда вещественный и инвариантный относительно поворотов вокруг начала координат.

Реальное растровое изображение $f_{n, m}$ имеет конечные размеры: $A \leq x \leq B$, $C \leq y \leq D$ и состоит из отдельных пикселей, расположенных с некоторым шагом в узлах прямоугольной сетки. В этом случае для перехода в частотную плоскость применяется двойное дискретное преобразование Фурье (ДДПФ):

$$F_{p, q} = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f_{n, m} \exp\left(-2\pi i \left[\frac{pn}{N} + \frac{qm}{M}\right]\right), \quad (3.13)$$

$$f_{n, m} = \sum_{p=0}^{N-1} \sum_{q=0}^{M-1} F_{p, q} \exp\left(2\pi i \left[\frac{pn}{N} + \frac{qm}{M}\right]\right). \quad (3.14)$$

Здесь N — число пикселей по x , M — число пикселей по y . Величины $F_{p, q}$, называемые коэффициентами ДДПФ, вычисляются следующим образом: вначале в (3.13) производится суммирование по n (по строкам), потом полученный числовой массив суммируется по m (по столбцам), можно и наоборот. Разработан также алгоритм, сводящий двумерное дискретное преобразование Фурье к одномерному.

Коэффициент $F_{0, 0}$ — это среднее значение яркости изображения. Общее число комплексных коэффициентов $F_{p, q}$ равно $N \cdot M$, однако часть из них связана между собой. Если, например, N — четное число, то $F_{N-p, q} = F_{p, q}^*$, т.е. комплексно-сопряженное, так что число независимых коэффициентов равно $N \cdot M/4$.

Для вычисления ДДПФ согласно (3.13) необходимо выполнить $N^2 \cdot M^2$ операций с комплексными числами, что требует значительных затрат времени. Если $N = M = 512$, то общее число операций

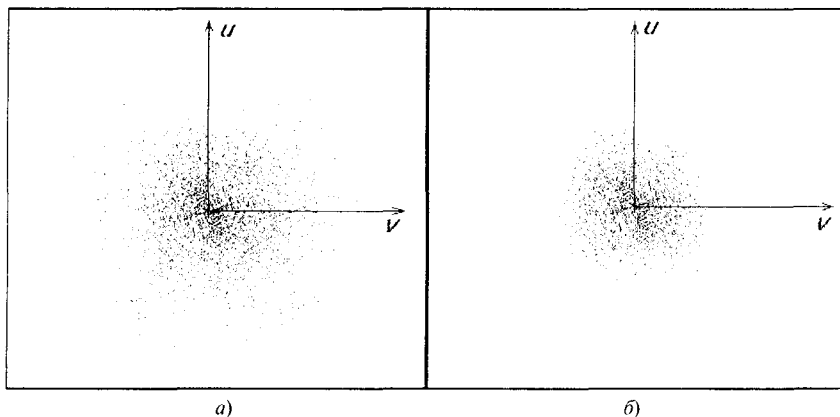


Рис. 3.11. Модуль спектра изображения на рис. 3.1 до (а) и после (б) фильтрации низко частотным фильтром

$512^4 \approx 6,8 \cdot 10^{10}$. Для ускорения вычислений разработаны алгоритмы быстрого преобразования Фурье (БПФ), общее число операций которых оценивается как $N \log N \cdot M \log M$, что для приведенного примера составляет $2 \cdot 10^7$ операций. Таким образом, вычисления ускоряются более чем в 3000 раз.

Глобальная линейная фильтрация изображений в частотной плоскости сводится к умножению $F_{p,q}$ на частотный коэффициент передачи $K_{p,q}$. Пример модуля ДДПФ и результат его умножения модуля спектра на коэффициент передачи вида $K_{p,q} = \exp(-|p + q|/a)$ показан на рис. 3.11.

3.3.2. Линейная фильтрация в частотной плоскости

При линейной фильтрации изображений в частотной плоскости требуется умножить спектр пространственных частот на частотный коэффициент передачи согласно (3.12) и выполнить два двумерных преобразования Фурье с использованием алгоритма БПФ — прямое и обратное. Преобразование Фурье осуществляется от всего изображения целиком, спектр $F(u, v)$ хранит информацию обо всем изображении целиком, поэтому можно говорить о глобальной фильтрации. В зависимости от выбора коэффициента передачи $K(u, v)$ можно выделять изображение на фоне помех, улучшать и ухудшать его резкость, усиливать контуры объектов на изображении и т.п.

Типичной задачей глобальной фильтрации, не характерной, впрочем, для цифровой обработки космических изображений земной поверхности, является восстановление расфокусированных изображе-

ний. Спектр такого изображения $F_1(u, v) = F(u, v) \cdot K(u, v)$, где $F(u, v)$ — спектр исходного изображения, $K(u, v)$ — коэффициент передачи оптической системы, соответствующей ФРТ на рис. 3.10. Примеры расфокусированных изображений приведены на рис. 3.12. Чтобы определить $F(u, v)$ по известным $F_1(u, v)$ и $K(u, v)$, необходимо умножить

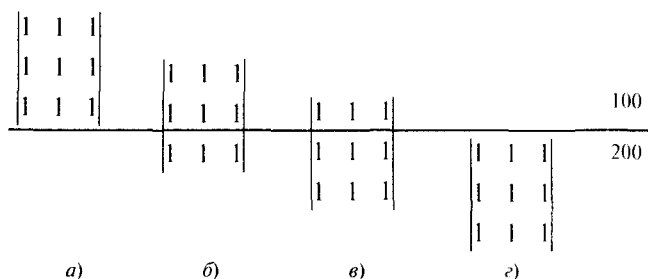


Рис. 3.12. Обработка перепада яркости маской H_2

$F_1(u, v)$ на $K_1(u, v) = 1/K(u, v)$, т.е. выполнить инверсную фильтрацию. Эта задача относится к числу обратных некорректных задач математической физики. Во-первых, решения могут не существовать. Во-вторых, если решение существует, то оно может быть не единственным. Заметим, что функция $K(u, v)$ в некоторых точках может быть равна нулю, тогда решение обращается в бесконечность. В-третьих, решение может быть неустойчивым, т.е. небольшие вариации исходных данных могут привести к существенным изменениям решения. Выходом из положения является регуляризация решения. Коэффициент передачи записывается в форме

$$K_1(u, v) = K(u, v) / [K^2(u, v) + \alpha \cdot g(u, v)]. \quad (3.15)$$

Здесь $g(u, v)$ — стабилизирующая функция, например $g(u, v) = u^2 + v^2$, α — параметр регуляризации. Подбором α обычно удается достаточно качественно восстановить расфокусированное изображение.

Фильтрация в частотной плоскости — эффективное средство для восстановления изображений, искаженных аддитивным шумом. Пусть изображение $s(x, y)$ наблюдается на фоне некоррелированного с $s(x, y)$ шума $n(x, y)$:

$$f(x, y) = s(x, y) + n(x, y).$$

Предполагается, что $f(x, y)$, $s(x, y)$ и $n(x, y)$ — однородные и изотропные гауссовские случайные поля с функциями корреляции $R_f(r)$,

$R_s(r)$, $R_n(r)$ соответственно, с нулевыми средними и ограниченными дисперсиями. При этом $R_s(r) = \langle s(x, y) s(x_1, y_1) \rangle$, остальные функции корреляции определены аналогичным образом; аргумент $r = \sqrt{(x - x_1)^2 + (y - y_1)^2}$. Из математической статистики известно, что если и полезный сигнал (в нашем случае это изображение), и шум имеют нормальный (гауссовский) закон распределения, то самый лучший эффект фильтрации, по крайней мере с точки зрения минимума среднего квадрата ошибки, обеспечивает линейный фильтр. Для выделения изображения на фоне шума используем линейное преобразование

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') h_1(x - x', y - y') dx' dy'.$$

Требуется найти ФРТ $h_1(x, y)$, обеспечивающую минимум среднего квадрата ошибки фильтрации $\epsilon^2 = \langle |g(x, y) - s(x, y)|^2 \rangle$. Находим, что

$$\begin{aligned} \epsilon^2 &= \left\langle \left[s(x, y) - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') h_1(x - x', y - y') dx' dy' \right]^2 \right\rangle = \\ &= R_s(r) - 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_s(r') h_1(x - x', y - y') dx' dy' + \\ &+ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{ R_s(r_1) + R_n(r_1) \} h_1(x - x', y - y') h_1(x_1 - x'_1, y_1 - y'_1) dx' dy' dx'_1 dy'_1, \end{aligned}$$

где $r' = \sqrt{(x' - x)^2 + (y' - y)^2}$, $r_1 = \sqrt{(x' - x'_1)^2 + (y' - y'_1)^2}$; учтено, что $\langle s(x, y) n(x_1, y_1) \rangle = 0$. Пусть h_1 мало отличается от оптимального значения h_1^* , т.е. $h_1 = h_1^* + \alpha \cdot \beta(x, y)$, здесь α — неопределенный множитель Лагранжа, $\beta(x, y)$ — некоторая функция. Необходимое условие минимума состоит в том, что $d\epsilon/d\alpha = 0$ при $\alpha = 0$. Таким образом,

$$\begin{aligned} & - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \beta(x - x', y - y') R_s(r') dx' dy' + \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \beta(x - x', y - y') R_f(r_1) h_1^*(x' - x'_1, y' - y'_1) dx' dy' dx'_1 dy'_1 = 0. \end{aligned}$$

Это равенство должно выполняться при произвольной $\beta(x - x', y - y')$. Откуда

$$-R_s(r') + \int_{-\infty}^{\infty} \int R_f(r_1) h_1^*(x' - x_1', y' - y_1') dx_1' dy_1' = 0. \quad (3.16)$$

Перейдем от функций корреляции к спектру мощности $G(u, v)$ и от ФРТ к частотному коэффициенту передачи $K(u, v)$. Спектр мощности связан с функцией корреляции преобразованием Фурье:

$$G(u, v) = \int_{-\infty}^{\infty} \int R(r) \exp(-2\pi i[u(x - x_1) + v(y - y_1)]) dx dy, \quad (3.17)$$

$$R(r) = \int_{-\infty}^{\infty} \int G(u, v) \exp(2\pi i[u(x - x_1) + v(y - y_1)]) du dv. \quad (3.17')$$

Подстановка (3.11') и (3.14) при $G_f(u, v) = G_s(u, v) + G_n(u, v)$ в (3.16) дает

$$-G_s(u, v) + K^*(u, v)[G_s(u, v) + G_n(u, v)] = 0,$$

откуда оптимальный коэффициент передачи

$$K^*(u, v) = \frac{G_s(u, v)}{G_s(u, v) + G_n(u, v)}. \quad (3.18)$$

Коэффициент передачи (3.18) обеспечивает минимальный средний квадрат ошибки фильтрации при выделении одного случайного поля (сигнала) на фоне другого случайного поля (шума) при условии, что оба поля — однородные и изотропные с гауссовским законом распределения значений яркости. Фильтр с таким коэффициентом передачи (3.18) называется фильтром Винера.

Когда изображение вначале искажено из-за воздействия некоторого линейного оператора, а затем на него наложен шум, можно объединить функции фильтра Винера и инверсного фильтра. В этом случае коэффициент передачи равен

$$K_2(u, v) = 1/[1 + K_1^2(u, v)/K^*(u, v)],$$

где $K_1(u, v)$ определяется выражением (3.15) при $\alpha = 0$. Гомоморфный (обобщенный) линейный фильтр имеет коэффициент передачи

$$K_3(u, v) = K_1(u, v) / \sqrt{1 + K_1^2(u, v)/K^*(u, v)}.$$

3.3.3. Линейная локальная фильтрация

На практике глобальная фильтрация применяется редко. Чаще используют локальную фильтрацию, когда интегрирование и усреднение проводятся не по всей области определения x и y , а по сравнительно небольшой окрестности каждой точки изображения. При этом функция рассеяния точки имеет ограниченные размеры. Достоинством такого подхода является хорошее быстроедействие.

При обработке растровых изображений, которые состоят из отдельных пикселей, интегрирование заменяют суммированием. Линейное преобразование в случае локальной фильтрации принимает вид

$$g_{ij} = \sum_D a_{kl} f_{i+k, j+l},$$

суммирование ведется по некоторой окрестности D точки (i, j) ; a_{kl} — значения ФРТ в этой окрестности. Яркости пикселей f в этой точке и в ее окрестности умножаются на коэффициенты a_{kl} , преобразованная яркость (i, j) -го пикселя есть сумма этих произведений. Обычно набор коэффициентов a_{kl} представляют в виде прямоугольной матрицы (маски), например размерности 3×3 :

$$\mathbf{H} = \frac{1}{K} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Элементы матрицы удовлетворяют условию пространственной инвариантности, поэтому $a_{11} = a_{13} = a_{31} = a_{33}$, $a_{12} = a_{21} = a_{23} = a_{32}$.

Фильтрация осуществляется перемещением слева направо (или сверху вниз) маски на один пиксел. При каждом положении апертуры производятся упомянутые операции: перемножение весовых множителей a_{kl} с соответствующими значениями яркостей исходного изображения и суммирование произведений. Полученное значение присваивается центральному (i, j) -му пикселу, которое обычно делится на заранее заданное число K (нормирующий множитель). Маска содержит нечетное число строк и столбцов, чтобы центральный элемент определялся однозначно.

Рассмотрим некоторые фильтры, сглаживающие шум. Пусть маска размером 3×3 имеет вид

$$H_1 = \frac{1}{10} \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{vmatrix}.$$

Тогда после фильтрации яркость (i, j) -го пиксела

$$g_{i,j} = 1/10(f_{i-1,j-1} + f_{i-1,j} + f_{i,j-1} + f_{i,j} + 2f_{i,j} + f_{i,j+1} + f_{i+1,j-1} + f_{i+1,j} + f_{i+1,j+1}).$$

Хотя коэффициенты a_{kl} можно выбрать из среднеквадратического или иного условия близости не искаженного шумом $s_{i,j}$ и преобразованного $g_{i,j}$ изображений, обычно их задают эвристически. Приведем еще некоторые матрицы шумоподавляющих фильтров:

$$H_2 = \frac{1}{9} \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix}, \quad H_3 = \frac{1}{16} \begin{vmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{vmatrix}, \quad H_4 = \frac{1}{14} \begin{vmatrix} 2 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 2 \end{vmatrix}.$$

У фильтров H_1-H_4 нормирующие множители K подобраны так, чтобы не происходило изменения средней яркости обработанного изображения. Наряду с масками 3×3 используются маски большей размерности, например 5×5 , 7×7 и т. п. В отличие от фильтра H_2 , у фильтров H_1 , H_3 , H_4 весовые коэффициенты на пересечении главных диагоналей матрицы больше, чем коэффициенты, стоящие на периферии. Фильтры H_1 , H_3 , H_4 дают более плавное изменение яркости по изображению, чем H_2 .

Пусть отсчеты неискаженного изображения мало меняются в пределах маски, а отсчеты аддитивного шума случайны и независимы или слабо зависимы со статистической точки зрения. В этом случае механизм подавления шума с использованием приведенных фильтров состоит в том, что при суммировании шумы компенсируют друг друга. Эта компенсация будет происходить тем успешнее, чем большее число членов в сумме, т.е. чем больше размер (апертура) маски. Пусть, например, используется маска $N \times N$, в пределах ее неискаженное изображение имеет постоянную яркость f , шум — аддитивный, с независимыми значениями отсчетов $n_{k,m}$, средним значением $\mu = 0$ и дисперсией σ^2 в пределах маски (такой шум называют белым). Отношение квадрата яркости (i, j) -го пиксела к дисперсии шума, т.е. отношение сигнал-шум, равно f^2/σ^2 .

Рассмотрим, например, маску типа H_2 :

$$g_{i,j} = \frac{1}{N^2} \sum_{k=1}^N \sum_{m=1}^N (f + n_{k,m}) = f + \frac{1}{N^2} \sum_{k=1}^N \sum_{m=1}^N n_{k,m}.$$

Средний квадрат яркости равен f^2 , средний квадрат интенсивности шума

$$\langle n_{i,j}^2 \rangle = \frac{1}{N^4} \sum_{k=1}^N \sum_{m=1}^N \langle n_{k,m}^2 \rangle + \frac{1}{N^4} \sum_{k=1}^N \sum_{m=1}^N \sum_{q=1}^N \sum_{p=1}^N \langle n_{k,m} n_{q,p} \rangle.$$

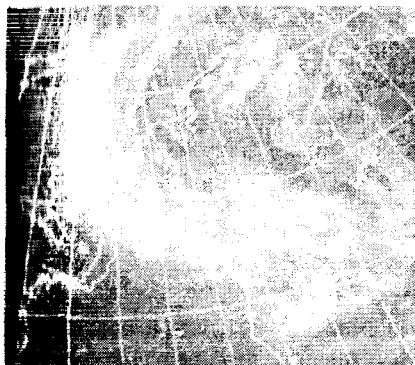
Двойная сумма, равная σ^2/N^2 , соответствует $k=p, m=q$. Четырехкратная сумма равна нулю, так как отсчеты шума при $k \neq p, m \neq q$ независимы: $\langle n_{k,m} n_{q,p} \rangle = 0$. Таким образом, в результате фильтрации отношение сигнал-шум становится равным $N^2 f^2 / \sigma^2$, т.е. возрастает пропорционально площади маски. Отношение яркости (i,j) -го пиксела неискаженного изображения к среднеквадратическому отклонению шума возрастает пропорционально N^2 . Следовательно, использование маски 3×3 в среднем повышает отношение сигнала к шуму в 9 раз.

При импульсной помехе механизм подавления состоит в том, что импульс «расплывается» и становится малозаметным на общем фоне.

Однако часто в пределах апертуры значения полезного изображения все же изменяются заметным образом. Это бывает, в частности, когда в пределы маски попадают контуры. С физической точки зрения все H_1 – H_4 являются *фильтрами нижних частот* (усредняющими фильтрами), подавляющими высокочастотные гармоники и шума, и неискаженного изображения. Это приводит не только к ослаблению шума, но и к размыванию контуров на изображении. Пусть, например, на изображении, обрабатываемом маской H_2 , имеется перепад яркости от 100 к 200 (см. рис. 3.12).

Если маска находится в положении *a*, когда в ее пределы попадают только значения яркости $f=100$, то $g_{i,j} = 9 \cdot 100/9 = 100$. Если маска находится в положении *г*, то $g_{i,j} = 9 \cdot 200/9 = 200$. Таким образом, в этих случаях яркость не изменяется.

Если же маска занимает положение *б*, то $g_{i,j} = (6 \cdot 100 + 3 \cdot 200)/9 = 133$. Для положения маски *в* имеем $g_{i,j} = (3 \cdot 100 + 6 \cdot 200)/9 = 167$. Следовательно, резкий профиль перепада яркости 100–200 стал сглаженным: 100–133–166–200. Результат применения фильтра H_2 с масками размерами 3×3 и 7×7 можно видеть на рис. 3.13 (оригинал приведен на рис. 3.14).

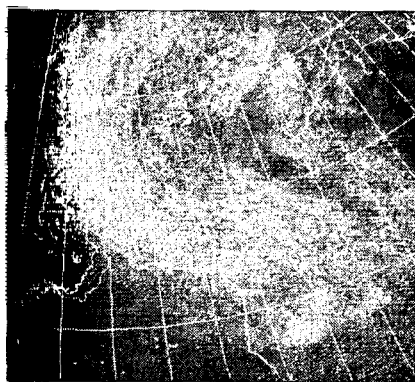


Маска 3×3

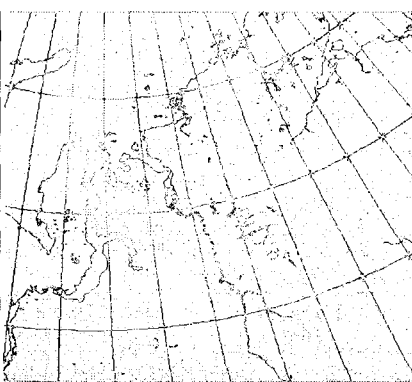


Маска 7×7

Рис. 3.13. Обработка изображения сглаживающим фильтром



Исходное изображение



Результат применения H_{13}

Рис. 3.14. Обработка оператором Лапласа

Рассмотренная фильтрация, называемая нерекурсивной, характеризовалась тем, что выходные значения фильтра g определялись только через входные значения фильтра f . Фильтрация, при которой выходные значения g определяются не только через входные значения f , но и через соответствующие выходные значения, называется рекурсивной.

При рекурсивной фильтрации можно сохранять те же значения весовых множителей, что и приведенные, существуют рекурсивные фильтры со специально подобранными множителями. Элементы входного изображения в пределах окна изменяются и примут вид

$$XY = \begin{vmatrix} g_{i-1, j-1} & f_{i-1, j} & f_{i-1, j+1} \\ g_{i, j-1} & f_{i, j} & f_{i, j+1} \\ g_{i+1, j-1} & f_{i+1, j} & f_{i+1, j+1} \end{vmatrix}.$$

Рассмотрим динамику изменения весовых множителей на примере рекурсивного фильтра первого рода с линейной маской размером в три пиксела. Пусть одномерный низкочастотный шумоподавляющий фильтр имеет весовые множители вида $\mathbf{H}_1 = |1, 1, 1|/3$.

При сдвиге апертуры фильтра на один пиксел вправо значения выхода фильтра $g_2 = (g_1 + f_2 + f_3)/3$. Подставляя сюда значение $g_1 = (f_0 + f_1 + f_2)/3$, можно получить, что весовые и нормирующие множители изменились и приняли вид $\mathbf{H}_2 = |1, 1, 4, 3|/9$. Повторяя проделанное, находим $\mathbf{H}_3 = |1, 1, 4, 12, 9|/27$, $\mathbf{H}_4 = |1, 1, 4, 12, 36, 27|/81$.

Если исходные весовые и нормирующие множители низкочастотного фильтра имеют вид $\mathbf{H}_1 = |1, 2, 1|/4$, то аналогично предыдущему случаю можно получить, что $\mathbf{H}_2 = |1, 2, 9, 4|/16$, $\mathbf{H}_3 = |1, 2, 9, 36, 16|/64$ и $\mathbf{H}_4 = |1, 2, 9, 36, 144, 64|/256$. Таким образом, весовые и нормирующие множители рекурсивного фильтра зависят от местоположения маски. рекурсивный локальный фильтр позволяет учитывать все входные значения фильтруемого изображения, т.е. приближается по своему действию к глобальному фильтру. Рекурсивный фильтр, вообще говоря, не является пространственно-инвариантным.

Если изображение описывается авторегрессионной моделью, для выделения его из шума можно использовать двумерный фильтр Калмана (ФК), по существу, являющийся вариантом фильтра Винера (ФВ). Фильтр Винера оптимален при всех допустимых спектрах сигнала и шума, описываемых однородными и изотропными гауссовскими случайными полями. Как и ФВ, ФК обеспечивает минимальный средний квадрат ошибки фильтрации изображения на фоне шума, но для существенно более узкого класса случайных полей. Он оптимален, когда сигнал имеет экспоненциальную функцию корреляции и спектр мощности вида $G_s(u, v) = 1/[1 + a(u^2 + v^2)]$. При этом спектр мощности шума $G_n(u, v) = \text{const}$ (такой шум называют белым).

Один из вариантов фильтра Калмана предложил А. Хабиби. Для трехточечной авторегрессионной модели оценка яркости (i, j) -го пиксела, по Хабиби, записывается как

$$g_{i,j} = \rho_1 g_{i-1,j} + \rho_2 g_{i,j-1} - (\rho_1 \rho_2 - F v_{i,j}) g_{i-1,j-1} + v_{i,j} f_{i,j},$$

где ρ_1 и ρ_2 — коэффициенты корреляции между соседними элементами по строке и столбцу, $F, v_{i,j}$ — коэффициенты, зависящие от дисперсии порождающего процесса, дисперсии шума и ρ_1, ρ_2 . Очевидно, что этот рекурсивный фильтр. Фильтр Винера определен в частотной плоскости, а фильтр Калмана — в плоскости координат. Фильтр Винера относит

к числу некаузальных (см. рис. 2.10, *г*), поскольку для вычисления преобразования Фурье используется вся плоскость координат. Фильтр Калмана — каузальный, поскольку для формирования оценки сигнала использует область, показанную на рис. 2.10, *б*. Отметим, что требование каузальности вступает в противоречие с условием пространственной инвариантности. Фильтр Винера обеспечивает несколько лучшее качество фильтрации, так как использует все изображение целиком, а ФК — только текущий и предыдущие пиксели. Важнейшей особенностью фильтра Калмана является быстроедействие, он обеспечивает выделение изображения на фоне шума в реальном времени. При обработке изображений Земли, получаемых со спутников, ФК применяется редко, так как работа в реальном времени, как правило, не требуется.

Линейные фильтры могут не только подавлять шум, но и подчеркивать перепады яркости и контуров. Выделение вертикальных перепадов осуществляется дифференцированием по строкам, горизонтальных — по столбцам. Дифференцирование производится в цифровой форме:

$$g_{i,k} = \Delta f_{i,k} / \Delta x = f_{i,k} - f_{i,k-1} \text{ (по строкам);}$$

$$g_{i,k} = \Delta f_{i,k} / \Delta y = f_{i,k} - f_{i-1,k} \text{ (по столбцам).}$$

Здесь $\Delta x = 1$ — приращение вдоль строки, равное 1 пикселу, $\Delta y = 1$ — приращение вдоль столбца, также равное 1 пикселу.

Выделение перепадов по диагонали можно получить, вычисляя разности уровней диагональных пар элементов. Для этого используют следующие наборы весовых множителей, реализующих двумерное дифференцирование:

$$\begin{array}{cccc} \begin{array}{c} \text{север} \\ \mathbf{H}_5 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{bmatrix} \end{array} & \begin{array}{c} \text{северо-восток} \\ \mathbf{H}_6 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{bmatrix} \end{array} & \begin{array}{c} \text{восток} \\ \mathbf{H}_7 = \begin{bmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{bmatrix} \end{array} & \begin{array}{c} \text{юго-восток} \\ \mathbf{H}_8 = \begin{bmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \\ \\ \begin{array}{c} \text{юг} \\ \mathbf{H}_9 = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} & \begin{array}{c} \text{юго-запад} \\ \mathbf{H}_{10} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} & \begin{array}{c} \text{запад} \\ \mathbf{H}_{11} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{bmatrix} \end{array} & \begin{array}{c} \text{северо-запад} \\ \mathbf{H}_{12} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{bmatrix} \end{array} \end{array}$$

Название географических направлений говорит об ориентации склона перепада, вызывающего максимальный отклик фильтра. Сумма весовых множителей масок равна нулю, поэтому на участках изображения с постоянной яркостью эти фильтры дают нулевой отклик.

Горизонтальный перепад можно также выделить путем вычисления приращения разности яркостей пикселей вдоль строки, что равноценно вычислению второй производной по направлению (оператор Лапласа):

$$g_{i,k} = \Delta^2 f_{i,k} / \Delta x^2 = [f_{i,k} - f_{i,k-1}] - [f_{i,k+1} - f_{i,k}] = 2f_{i,k} - f_{i,k-1} - f_{i,k+1}.$$

Это соответствует одномерной маске $\mathbf{H} = [-1 \ 2 \ -1]$, сумма весовых множителей равна нулю. Таким же образом можно искать перепады по вертикали и по диагонали. Для выделения перепадов без учета их ориентации используются двумерные операторы Лапласа:

$$\mathbf{H}_{13} = \begin{vmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{vmatrix}, \quad \mathbf{H}_{14} = \begin{vmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{vmatrix}, \quad \mathbf{H}_{15} = \begin{vmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{vmatrix}.$$

Здесь сумма весовых множителей также равна нулю. На рис. 3.14 приведены исходное изображение и результат применения оператора Лапласа \mathbf{H}_{13} .

Операторы Лапласа реагируют не только на ступенчатый и «крышеобразный» перепады яркости, но и на изолированные точки, тонкие линии, их концы и острые углы объектов: на линию — в 2 раза ярче, чем ступенчатый перепад, на конец линии — в 3 раза ярче, а на точку — в 4 раза. Поскольку оператор Лапласа выделяет в основном не структурированные элементы, он чувствителен к шуму. Хотя подчеркивание перепадов с помощью оператора Лапласа происходит без учета их ориентации, этот оператор не инвариантен к ориентации перепадов: например, его отклик на наклонный перепад в диагональном направлении почти вдвое больше, чем в горизонтальном и вертикальном.

С физической точки зрения фильтры \mathbf{H}_5 – \mathbf{H}_{15} представляют собой фильтры верхних частот: они выделяют высокочастотные составляющие неискаженного изображения, ответственные за перепады яркости и контуры, и подавляют «постоянную составляющую». Однако при их использовании уровень шума на изображении возрастает.

Фильтры для выделения перепадов и границ, как и фильтры \mathbf{H}_1 – \mathbf{H}_4 для подавления шума, могут быть рекурсивными.

Изображение с подчеркнутыми границами (контурами) субъективно воспринимается лучше, чем оригинал. При использовании оператора Лапласа для этих целей применяют три типовых набора весовых множителей:

$$\mathbf{H}_{16} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \mathbf{H}_{17} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \mathbf{H}_{18} = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix},$$

отличающиеся от фильтров \mathbf{H}_{13} – \mathbf{H}_{15} тем, что к центральному элементу матрицы прибавлена 1, т.е. при фильтрации исходное изображение накладывается на контур.

Для авторегрессионной модели изображения существуют эффективные алгоритмы подчеркивания контуров и перепадов яркости. При трехточечной схеме авторегрессии весовые множители фильтра 3×3 имеют вид

$$\mathbf{H} = \begin{bmatrix} \rho_1 \rho_2 & -\rho_1 (1 + \rho_2^2) & \rho_1 \rho_2 \\ -\rho_2 (1 + \rho_1^2) & (1 + \rho_1^2)(1 + \rho_2^2) & -\rho_2 (1 + \rho_1^2) \\ \rho_1 \rho_2 & -\rho_1 (1 + \rho_2^2) & \rho_1 \rho_2 \end{bmatrix},$$

где ρ_1, ρ_2 — по-прежнему коэффициенты корреляции между соседними элементами по строке/столбцу. При $\rho_1 = \rho_2 = 0$ все весовые множители равны нулю, кроме центрального, который равен 1, и маска не изменяет исходного изображения. При $\rho_1 = \rho_2 = 1$ маска переходит в оператор Лапласа \mathbf{H}_{15} .

Для повышения помехоустойчивости при выделении и наложении контуров, для устранения эффекта размывания контуров при подавлении шума следует переходить к нелинейной обработке.

Как указывалось, если и полезный сигнал (в нашем случае это изображение), и шум имеют нормальный закон распределения, то самый лучший эффект фильтрации, с точки зрения получения минимума среднего квадрата ошибки, обеспечивает линейный фильтр. Реальные изображения по своим статистическим характеристикам часто не соответствуют нормальному закону распределения, и потому оптимальный алгоритм следует, вообще говоря, искать среди нелинейных алгоритмов фильтрации.

3.4. НЕЛИНЕЙНАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ

3.4.1. Медианная фильтрация

Для подавления шума, построенного из эвристических соображений, применяют нелинейный медианный фильтр. Медианная филь-

рации, а также алгоритмы экстремальной фильтрации, использующие значения минимума и максимума текущей окрестности, относятся к так называемым ранговым методам.

В общем случае ранговая фильтрация описывается функцией членов вариационного ряда V_1, \dots, V_n значений элементов текущей окрестности. Вариационным рядом V_1, \dots, V_n выборки f_1, \dots, f_n называют упорядоченную по неубыванию последовательность элементов выборки, т.е. $V_1 = \min(f_1, \dots, f_n)$, $V_n = \max(f_1, \dots, f_n)$ и т.д. В нашем случае выборка берется по (k, l) -окрестности точки с координатами (i, j) . Окрестность, как и ранее, задается некоторой маской.

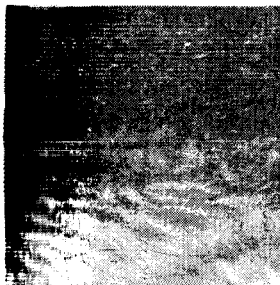
Максимальная фильтрация $g_{i,j} = \max\{f_{i+k,j+l}\}$ обладает следующими свойствами: независимость от ориентации импульсов, сохранение резкости, независимость от отрицательных выбросов, что позволяет устранять помехи изображения, обусловленные дефектными пикселями, строками и/или столбцами. Присвоение (i, j) -му пикселу максимального по окрестности значения — операция нелинейная.

Минимальная фильтрация $g_{i,j} = \min\{f_{i+k,j+l}\}$ также не зависит от ориентации импульсов и сохраняет резкость. Кроме этого, она нечувствительна к положительным выбросам, сужает размеры импульсов, делает линии более тонкими, тем самым частично исправляя искажения типа дефокусировки. Присвоение (i, j) -му пикселу минимального по окрестности значения яркости — операция также нелинейная.

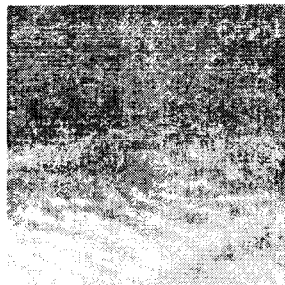
Медианная фильтрация $g_{i,j} = \text{Me}\{f_{i+k,j+l}\}$, сочетая многие полезные свойства максимального и минимального фильтров, нечувствительна как к отрицательным, так и к положительным выбросам. При медианной фильтрации (i, j) -му пикселу присваивается медианное значение яркости, т.е. такое значение, частота которого равна 0,5. Пусть, например, используется маска 3×3 , в пределы которой наряду с более или менее равномерным фоном попал шумовой выброс, этот выброс пришелся на центральный элемент маски:

$$\begin{vmatrix} 68 & 79 & 74 \\ 63 & 212 & 83 \\ 66 & 71 & 89 \end{vmatrix}.$$

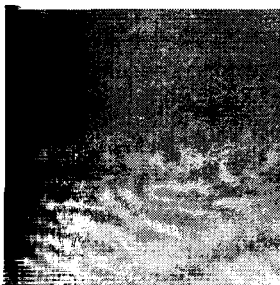
Вариационный ряд имеет вид 63, 66, 68, 71, 74, 79, 83, 89, 212. В данном случае медианное значение — пятое по счету (подчеркнуто), так как всего чисел в ряду 9. При этом значение 212, искаженное шумовым выбросом, заменяется на 74, выброс на изображении полностью подавлен. На рис. 3.15 приведены исходное изображение и результат применения



Исходное изображение



Медианный фильтр, окно 3×3



Медианный фильтр, окно 7×7



Линейный фильтр H_2 , окно 5×5

Рис. 3.15. Сравнение нелинейной и линейной фильтрации

Медианного фильтра, для сравнения показано действие линейного фильтра.

Медианный фильтр обладает некоторыми чертами линейного фильтра: $Me\{\alpha f_{ij}\} = \alpha Me\{f_{ij}\}$, $Me\{f_{ij} + c_0\} = c_0 + Me\{f_{ij}\}$, где α и c_0 — постоянные величины. Однако поскольку принцип суперпозиции не выполняется: $Me\{f_{ij} + s_{ij}\} \neq Me\{f_{ij}\} + Me\{s_{ij}\}$, то это нелинейный фильтр. Функция рассеяния точки для медианного фильтра есть нуль. При размерах окна $(2k+1)(2k+1)$ полностью подавляются помехи, состоящие не более чем из $2(k^2+k)$ пикселей, а также те, которые пересекаются не более чем k строками или k столбцами. При этом в точках фона не изменяется яркость. Разумеется, при медианной фильтрации может происходить искажение объекта на изображении, но только на границе или вблизи нее, если размеры объекта больше размеров маски. Фильтр обладает высокой эффективностью при подавлении импульсных помех, однако это качество достигается подбором размеров маски, когда известны минимальные размеры объектов и максимальные размеры искаженных помехой локальных областей. На случайный шум с нормальным законом распределения такой фильтр воздействует слабее, чем линейный усредняющий фильтр (приблизительно на 60% менее эффективно).

Вариационный ряд V_1, \dots, V_n текущей окрестности можно использовать для синтеза так называемых L -фильтров. В общем случае L -фильтр дает оценку $g_{i,j}$ полезного сигнала в виде некоторой функции от членов вариационного ряда $g_{i,j} = F(V_1, \dots, V_n)$. Как и при линейной фильтрации различают линейную L -фильтрацию (функция F — линейная), квадратичную (функция F — квадратичная) и т.д.

Задача синтеза L -фильтра решается так же, как и в случае линейной квадратичной и т.д. фильтрации. При этом необходимо вычислять начальные моментные функции членов вариационного ряда на основе заданных эталонных импульсов, которые выделяются на изображении.

3.4.2. Сигма-фильтр

Помимо ранговых алгоритмов фильтрации, при нелинейной обработке изображений применяют пороговые алгоритмы, из которых наиболее известен σ -фильтр.

Идея построения σ -фильтра основана на том, что закон распределения шума можно приближенно считать гауссовским. Часто шум обусловлен многими независимыми или слабо зависимыми факторами. Известно, например, что сумма небольшого числа (5–6) независимых равномерно распределенных слагаемых достаточно хорошо подчиняется гауссовскому закону распределения. Свойство нормализации закона распределения суммы независимых случайных величин (при условии равномерно малого вклада слагаемых в сумму) следует из центральной предельной теоремы теории вероятностей. Предположение о гауссовском характере закона распределения суммы выполняется тем лучше, чем большее число факторов ее обуславливают.

95,5% значений гауссовского распределения отклоняются от среднего значения меньше, чем 2σ . В соответствии с алгоритмом σ -фильтра оценивается среднее значение и среднеквадратическое отклонение яркостей в пределах достаточно большой маски; пиксели, имеющие яркость выше этого допуска, исключаются. Вместо исключенных пикселей подставляется среднее по окрестности, возможна интерполяция с использованием схемы авторегрессии. Выбранный допуск не настолько велик, чтобы исказить пиксели, принадлежащие крутым краям и мелким деталям, а также слабо искаженные шумом. σ -фильтр наиболее эффективен в борьбе с импульсными помехами в виде небольших точек. Вероятность того, что такая помеха попадет на контур, обычно невелика, поскольку при использовании σ -фильтра контуры практически не искажаются. В этом состоят преимущества σ -фильтра по сравнению с линейным, медианным, L -фильтрами. Однако такие

Преимущества заметнее всего при распределении шума на изображении, близком к нормальному (см. рис. 3.15). σ -фильтр может быть реализован также в рекурсивной форме.

При использовании обратногоградиентного фильтра вычисляется усредненное среднее с коэффициентами, обратно пропорциональными модулю градиента изображения в соответствующей точке. Основное преимущество обратногоградиентных фильтров — сглаживание однородных областей с сохранением границ между ними.

4.3. Нелинейные градиентные фильтры

Нелинейные фильтры, как и линейные, позволяют выделять контуры и перепады яркости. Отличие нелинейных алгоритмов состоит в том, что они используют нелинейные операторы дискретного дифференцирования. В фильтре Робертса применяется перемещающаяся по изображению маска 2×2 :

$$\begin{vmatrix} f_{i,j} & f_{i,j+1} \\ f_{i+1,j} & f_{i+1,j+1} \end{vmatrix},$$

дифференцирование производится с помощью одного из выражений:

$$g_{i,j} = \{(f_{i,j} - f_{i+1,j+1})^2 + (f_{i,j+1} - f_{i+1,j})^2\}^{1/2},$$

$$g_{i,j} = |f_{i,j} - f_{i+1,j+1}| + |f_{i,j+1} - f_{i+1,j}|.$$

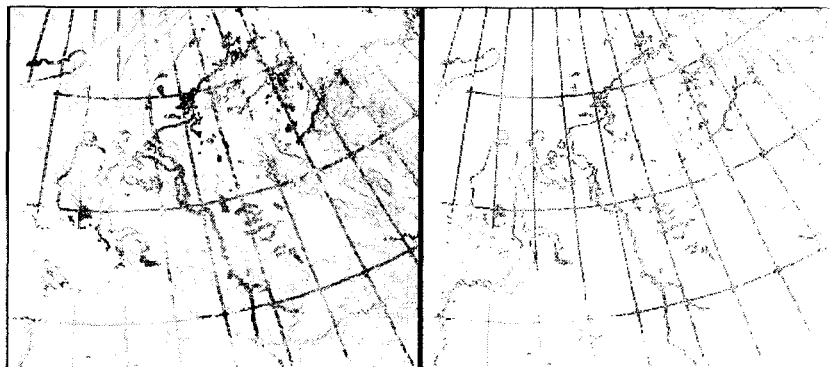
В фильтре Собела используется окно 3×3 :

$$\begin{vmatrix} f_{i-1,j-1} & f_{i-1,j} & f_{i-1,j+1} \\ f_{i,j-1} & f_{i,j} & f_{i,j+1} \\ f_{i+1,j-1} & f_{i+1,j} & f_{i+1,j+1} \end{vmatrix}.$$

Центральному (i, j) -му пикселу вместо $f_{i,j}$ присваивается значение яркости $g_{i,j} = (X^2 + Y^2)^{1/2}$ либо $g_{i,j} = |X| + |Y|$, где $X = (f_{i-1,j-1} + 2f_{i-1,j} + f_{i-1,j+1}) - (f_{i+1,j-1} + 2f_{i+1,j} + f_{i+1,j+1})$, $Y = (f_{i-1,j-1} + 2f_{i,j-1} + f_{i+1,j-1}) - (f_{i-1,j+1} + 2f_{i,j+1} + f_{i+1,j+1})$.

На рис. 3.16 показаны результаты применения фильтров Робертса и Собела к изображению, показанному на рис. 3.14.

При логарифмическом методе выделения контуров используют оператор Лапласа \mathbf{H}_{13} , причем вместо яркостей подставляют их логарифмы.



Фильтр Робертса

Фильтр Собела

Рис. 3.16. Обработка изображения нелинейными градиентными фильтрами

При сравнении логарифма яркости центрального пиксела, умноженного согласно H_{13} на 4, с логарифмами яркости четырех его соседей:

$$L = 4\log f_{i,j} - \log f_{i-1,j} - \log f_{i+1,j} - \log f_{i,j-1} - \log f_{i,j+1},$$

(i, j) -й пиксел относят к контуру или перепаду яркости, если его логарифм яркости превосходит среднее значение логарифма яркости четырех его ближайших соседей, т.е. когда $L > 0$. Выражение для L может быть переписано в виде

$$L = \log \frac{f_{i,j}^4}{f_{i-1,j} f_{i+1,j} f_{i,j-1} f_{i,j+1}}. \quad (3.19)$$

При этом необходимо предварительно принять меры, чтобы знаменатель был больше единицы либо, во всяком случае, больше нуля. Логарифмирование уменьшает диапазон возможных значений яркости, поэтому (3.19) следует применять там, где необходимо выделять незначительные перепады яркости («рябь» на изображении).

Пусть изображение искажено мультипликативной помехой, приводящей к ослаблению (усилению) яркости в отдельных участках. Причиной такого рода искажений при исследовании Земли из космоса могут быть неоднородные по поглощению участки атмосферы. Величина L нечувствительна к мультипликативным помехам, если в пределах окна яркости всех пикселов изменяются в одинаковое число раз: $m \cdot f_{i,j}$, $m \cdot f_{i-1,j}$, $m \cdot f_{i+1,j}$, $m \cdot f_{i,j-1}$, $m \cdot f_{i,j+1}$.

При функциональной фильтрации происходит вычисление обобщенного среднего. В самом общем виде под функциональной фильтрацией понимаются преобразования вида $z_{ij} = \alpha F^{-1}(1/n \sum F(f_{i+k, j+l}))$, где α — коэффициент, вводимый для сохранения средней яркости, n — число пикселей текущей области D .

До настоящего времени исследованы не все допустимые классы функции F . Некоторые виды таких преобразований ($F=f^2$, $F=f^{1/2}$) используются в системах технического зрения роботов. Достаточно общий и широкий класс функциональных фильтров образует обобщенное усреднение по Колмогорову (степенное среднее), что соответствует $F=f^p$. Особенностью функционального фильтра является то, что резкие границы изображения приобретают форму $F^{-1}(\mathbf{n})$, где \mathbf{n} — нормаль к границе, направленная в сторону больших значений. Например, класс линейных преобразований переводит прямоугольный импульс в трапециевидный. Класс обобщенно-усредняющих фильтров в качестве крайних случаев включает в себя экстремальные (т.е. минимальный и максимальный) фильтры. Минимальному фильтру соответствует $p \rightarrow 0$, максимальному $p \rightarrow \infty$.

Иногда в задачах обработки изображений может оказаться полезным применение функциональных фильтров типа степенного среднего с взаимно-обратными показателями: p и $1/p$. Эти фильтры позволяют целенаправленно изменять форму локальных минимумов или максимумов на изображении. Например, среднеквадратичный фильтр $F=f^2$ обостряет локальные максимумы и сглаживает локальные минимумы. В свою очередь, фильтр с обратным показателем $p = 1/2$, наоборот, обостряет локальные минимумы и сглаживает локальные максимумы.

4.4. Кепстральная обработка

Вернемся к задаче выделения изображений на фоне помех. Модель искаженного помехами изображения записывается в виде

$$f(x, y) = m(x, y) \cdot s(x, y) + n(x, y),$$

где $s(x, y)$ — неискаженное изображение, $m(x, y)$ — мультипликативная помеха, $n(x, y)$ — аддитивная помеха. Уже рассмотрены достаточно эффективные линейные и нелинейные методы устранения аддитивных помех на изображении. Труднее бороться с мультипликативными помехами, которые могут быть обусловлены, например, различной прозрачностью участков атмосферы. Если мультипликативная помеха

$m(x, y)$ постоянна в пределах объекта на изображении, то с успехом может быть применен логарифмический фильтр, работающий в плоскости координат. Пример такого фильтра, предназначенного для выделения контуров при наличии мультипликативной помехи, приведен в выражении (3.19).

Сложнее, когда помеха образует свертку с сигналом:

$$f(x, y) = \int\limits_{AC}^B \int\limits_C^D s(x', y') m(x, y, x', y') dx' dy', \quad (3.20)$$

изображение имеет конечные размеры: $A \leq x \leq B$; $C \leq y \leq D$. Если перейти в плоскость пространственных частот, то вместо (3.20) имеем

$$F(u, v) = S(u, v) \cdot M(u, v), \quad (3.21)$$

где $F(u, v)$, $S(u, v)$, $M(u, v)$ — преобразования Фурье от $f(x, y)$, $s(x, y)$, $m(x, y)$ соответственно, причем $F(u, v)$ и $S(u, v)$ — пространственные спектры искаженного и истинного изображения. Прологарифмировав (3.21), получим

$$\ln F(u, v) = \ln S(u, v) + \ln M(u, v).$$

Для вещественных функций $f(x, y)$ и $s(x, y)$ спектры, вообще говоря, являются комплексными:

$$F(u, v) = F_1(u, v) + iF_2(u, v) = |F(u, v)| \exp i\varphi(u, v),$$

где $F_1(u, v)$ и $F_2(u, v)$ — действительная и мнимая части,

$$\begin{aligned} |F(u, v)| &= \sqrt{F_1^2(u, v) + F_2^2(u, v)} \text{ — модуль;} \\ \varphi(u, v) &= \arcsin F_2(u, v)/F_1(u, v) \text{ — фазовый спектр.} \end{aligned} \quad (3.22)$$

Для изображения $s(x, y)$

$$S(u, v) = |S(u, v)| \exp i\mu(u, v);$$

если $m(x, y)$ описывает пространственно-инвариантное преобразование, то

$$M(u, v) = |M(u, v)|,$$

для пространственно-неинвариантных преобразований

$$M(u, v) = |M(u, v)| \exp i\psi(u, v).$$

Таким образом, для пространственно-инвариантного случая

$$\ln F(u, v) = \ln |S(u, v)| + \ln |M(u, v)| + i\varphi(u, v),$$

для пространственно-неинвариантного

$$\ln F(u, v) = \ln |S(u, v)| + \ln |M(u, v)| + i\varphi(u, v) + i\mu(u, v).$$

Кепстр, определяемый соотношением

$$K(x, y) = F^{-1} [\ln |F(u, v)| \exp i\mu(u, v)],$$

где F^{-1} обозначает обратное преобразование Фурье, является вещественной четной функцией (слово «кепстр» получено путем перестановки букв в слове «спектр»).

Используя кепстр и оценив фазовый спектр согласно (3.22), вместо исходного изображения $f(x, y)$ получаем преобразованное:

$$A(x, y) = F^{-1} [\ln |S(u, v)| \exp i\mu(u, v)] + F^{-1} [\ln |M(u, v)| \exp i\mu(u, v)]. \quad (3.23)$$

Важной проблемой при кепстральном анализе является восстановление фазового спектра. Теоретически при пространственно-инвариантном преобразовании (3.20) должно быть $\mu(u, v) = \varphi(u, v)$. Однако реально существует проблема восстановления фазы. Дело в том, что (3.22) позволяет определить фазовый спектр $\varphi(u, v)$ только на интервале $[-\pi, \pi]$. Если функция $\varphi(u, v)$ достигнет π , то произойдет скачок фазы на $-\pi$ вместо монотонного изменения. Разработаны различные методы восстановления фазы, которые используются в случае пространственно-неинвариантного преобразования.

Замечательной особенностью кепстра является то, что он согласно (3.23) переводит мультипликативную помеху, «запрятанную» в свертку, в аддитивную, методы борьбы с которой хорошо разработаны.

4.5. Ряд Вольтерра в нелинейной обработке изображений

К числу перспективных фильтров относятся фильтры, использующие ряд Вольтерра (3.6). При цифровой обработке вместо интегрирования используется суммирование

$$g_{i,j} = \sum_k \sum_l h_1(i,j) f_{i+k, j+l} + \sum_k \sum_l \sum_p \sum_q h_2(i,j, m, n) f_{i+k, j+l} f_{m+p, n+q}$$

в некоторой окрестности точек i, j и m, n , задаваемой в виде квадратной маски с нечетным числом строк и столбцов N . Ядро первого порядка задается в виде матрицы \mathbf{H} , похожей на матрицы \mathbf{H}_1 – \mathbf{H}_4 . Ядро второго порядка h_2 — это блочная матрица $N \times N$, составленная из простых матриц такой же размерности. Если окрестность задана маской 3×3 , то при вычислении второго члена ряда Вольтерра необходимо перемножить $9 \times 9 = 81$ значение яркости, каждое из таких произведений умножается на весовой коэффициент $h_2(i, j, m, n)$, число которых в принципе также равно 81. Результат суммирования упомянутых двух членов ряда Вольтерра присваивается (i, j) -му пикселу обработанного изображения. При выборе $h_1(i, j)$ и $h_2(i, j, m, n)$ учитываются условия симметрии — этот набор коэффициентов должен удовлетворять условию пространственной инвариантности, в частности быть инвариантным к произвольным поворотам вокруг точки (i, j) .

Как и в случае линейных фильтров \mathbf{H}_1 – \mathbf{H}_4 , в составе \mathbf{H} есть только три независимых коэффициента. Часть членов матрицы коэффициентов $h_2(i, j, m, n)$ также совпадает, например коэффициенты при $f_{ij} \cdot f_{mn}$ и $f_{mi} \cdot f_{jn}$ должны быть одинаковыми. Кроме того, из условия изотропности оператора Вольтерра следует, что матрица коэффициентов $h_2(i, j, m, n)$ не должна измениться, если исходное изображение повернуть на произвольный угол. На практике при прямоугольных окнах удается добиться инвариантности только при поворотах, кратных 90° (достаточно рассмотреть три последовательных поворота на 90°). Поэтому матрица содержит лишь 11 независимых коэффициентов.

Основная проблема синтеза фильтров Вольтерра — выбор ядер. В частности, их можно найти из условия минимума среднего квадрата ошибки фильтрации: $\min \langle [g(x, y) - s(x, y)]^2 \rangle$, однако, как и при линейной фильтрации, ядра чаще подбирают эвристически.

Фильтры, использующие ряд Вольтерра, позволяют эффективно подавлять помехи на изображении и одновременно контрастировать, выделять контуры.

В качестве примера рассмотрим фильтр для подавления шума на изображении при сохранении резкости (благодаря наложению контуров на отфильтрованное изображение) используются операторы первого и второго порядков, маска $h_1(i, j)$ представлена в виде таблицы 3×3 ; маска $h_2(i, j, m, n)$ — в виде таблицы 9×9 . При обработке по изображению

нию перемещается окно 3×3 . Пусть в него попали пиксели с яркостью f_{11}, f_{12}, f_{13} и т.д.:

f_{11}	f_{12}	f_{13}
f_{21}	f_{22}	f_{23}
f_{31}	f_{32}	f_{33}

Линейная фильтрация выполняется по обычному правилу. Нелинейный член подвергается следующим операциям. Значение яркости пиксела f_{11} , попавшее в окно 3×3 , умножается последовательно на все f_{ij} из этого окна с весами $h_2(i, j, m, n)$, взятыми из первой строки табл. 3.1. Так, произведение $f_{11} \cdot f_{11}$ умножается на $-0,0009$, произведение $f_{11} \cdot f_{12}$ — на $0,0001$ и т.д. Значение яркости пиксела f_{12} последовательно перемножается с весами $h_2(i, j, m, n)$, взятыми из второй строки. Процесс продолжается, пока не будет получено полностью все $9 \times 9 = 81$ произведение. Сумма произведений присваивается центральному элементу окна, т.е. f_{22} .

Матрица весовых коэффициентов $h_1(i, j)$

$$H = \begin{vmatrix} 2 & 0,25 & 2 \\ 0,25 & 1 & 0,25 \\ 2 & 0,25 & 2 \end{vmatrix}.$$

Т а б л и ц а 3.1. Таблица весовых коэффициентов $h_2(i, j, m, n)$

m \ n	m, n								
	11	12	13	21	22	23	31	32	33
11	-0,0009	0,0001	-0,0001	-0,0001	0,0002	-0,0001	-0,0001	-0,0001	-0,0002
12	0,0001	0,0005	-0,0001	0	0,0001	0	-0,0001	-0,0001	-0,0001
13	-0,0001	-0,0001	-0,0009	-0,0001	0,0002	0,0001	-0,0002	-0,0001	-0,0001
21	-0,0001	0	-0,0001	0,0005	0,0001	-0,0002	0,0001	0	-0,0001
22	0,0002	0,0001	0,0002	0,0001	0,0018	0,0001	0,0002	0,0001	0,0002
23	-0,0001	0	0,0001	-0,0002	0,0001	0,0005	-0,0001	0	-0,0001
31	-0,0001	-0,0001	-0,0002	0,0001	0,0002	-0,0001	-0,0009	-0,0001	-0,0001
32	-0,0001	-0,0001	-0,0001	0	0,0001	0	-0,0001	0,0005	0,0001
33	0,0002	-0,0001	-0,0001	-0,0001	0,0002	-0,0001	-0,0001	0,0001	-0,0009

Коэффициенты $h_1(i, j)$ подобраны так, чтобы подавлять шум в виде точек на изображении. Пусть такая точка имеет интенсивность 200 и наблюдается на фоне 100. В результате воздействия оператора первого порядка шумовой выброс интенсивностью в пределах 102–120 расплывается; контуры на изображении также расплываются. Оператор второго порядка выделяет контуры без заметного увеличения шумовых выбросов. Результат применения этого фильтра к изображению на рис. 3.14 можно видеть на рис. 3.17.

Процедуры нелинейной фильтрации богаче других по возможностям, более гибкие при изменениях характеристик полезного сигнала и шума. Выбор типа фильтра (линейного или нелинейного, вида нелинейного) зависит от специфики решаемой задачи, от наличия априорных сведений о свойствах полезного сигнала и шума, от требований к качеству изображения, от ресурсов вычислительной системы.

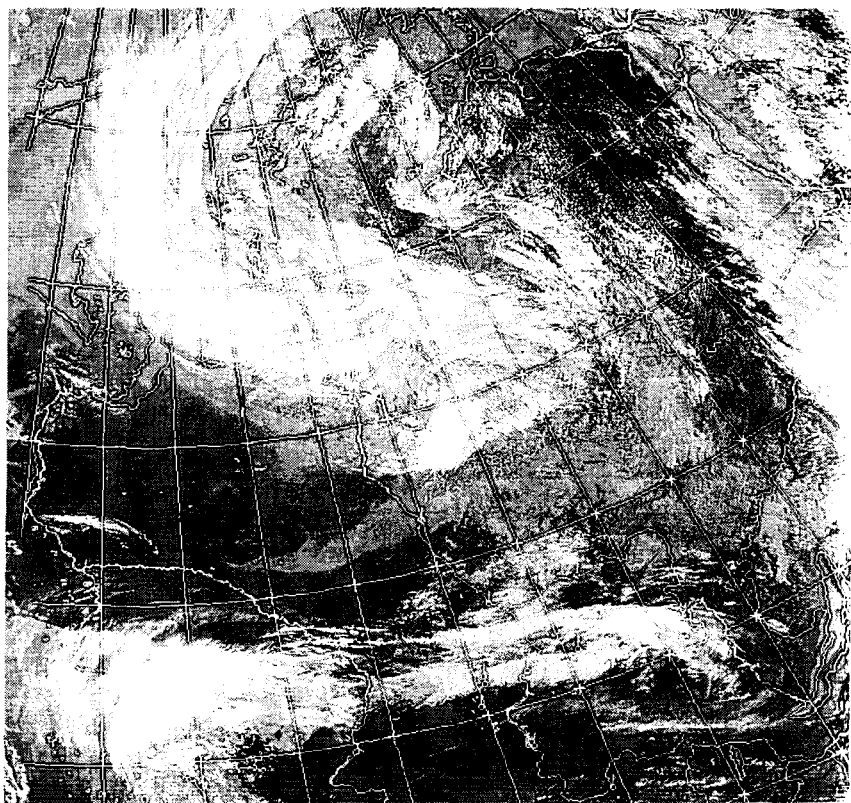


Рис. 3.17. Обработка изображения с помощью ряда Вольтерра

КОНТРОЛЬНЫЕ ЗАДАНИЯ

1. Каково минимальное число опорных точек, необходимых для геометрической коррекции изображения при использовании полиномов первой и третьей степени?
2. Напишите систему уравнений для определения коэффициентов полинома второй степени при геометрической коррекции изображения.
3. Показания датчика 4-го канала сканера AVHRR спутника NOAA равны 120. Какова температура объекта на поверхности Земли?
4. Спутник наблюдает некоторый район суши под углом 22° к надиру, оптическая толщина атмосферы $\tau = 0,12$. Во сколько раз атмосфера ослабляет излучение от объектов в этом районе?
5. Используя пакет прикладных программ для обработки изображений, выполните на ПЭВМ растяжку, нормализацию и эквализацию гистограммы.
6. На фоне изображения с постоянной яркостью $f = 120$ наблюдается точка размером в 1 пиксел с яркостью 10, а также в стороне линия шириной 1 пиксел с яркостью 200. Подсчитайте, что произойдет с этими объектами, если к изображению применить фильтр H_1 .
7. Как изменяются коэффициенты рекурсивного фильтра H_4 при перемещении маски по изображению?
8. На изображении наблюдается перепад яркости от 100 вверху к 200 внизу. Какова реакция фильтров H_5 , H_7 , H_9 на этот перепад? Отрицательным значениям яркости присваивать 0.
9. На изображении наблюдается перепад яркости от 100 вверху к 200 внизу. Какова реакция фильтра Лапласа H_{14} на этот перепад? Как реагирует оператор Лапласа на точку размером в 1 пиксел с яркостью 10 на фоне изображения с яркостью 100, а также на линию шириной 1 пиксел с яркостью 150? Рассмотрите варианты, когда отрицательным значениям яркости присваивается 0 и положительное значение по модулю.
10. На изображении наблюдается перепад яркости от 100 вверху к 200 внизу. Какова реакция медианного фильтра на этот перепад? Как реагирует медианный фильтр на точку размером в 1 пиксел с яркостью 10 на фоне изображения с яркостью 100, а также на линию шириной 1 пиксел с яркостью 150?
11. На изображении наблюдается перепад яркости от 100 вверху к 200 внизу. Какова реакция каждого из двух вариантов фильтра Робертса на этот перепад?
12. На изображении наблюдается перепад яркости от 100 вверху к 200 внизу. Какова реакция каждого из двух вариантов фильтра Собела на этот перепад?

ЛИТЕРАТУРА

Ишшин В.В. Анализ и обработка изображений: принципы и алгоритмы. М.: Машиностроение, 1995. 112 с.

Прэтт У. Цифровая обработка изображений: В 2-х кн. М.: Мир, 1987. 310 с.; 360 с.

Павлидис Т. Алгоритмы машинной графики и обработки изображений. М.: Радио и связь, 1986. 400 с.

Автоматизированная обработка изображений природных комплексов Сибири / Под ред. А.Л. Яншина и В.А. Соловьева. Новосибирск: Наука, 1988. 224 с.

Дистанционное зондирование: количественный подход. Пер. с англ. под ред. А.С. Алексеева. М.: Недра, 1983. 410 с.

Ярославский Л.П. Цифровая обработка сигналов в оптике и голографии. М.: Радио и связь, 1987. 296 с.

Василенко Г.И., Тараторин А.М. Восстановление изображений. М.: Радио и связь, 1986. 302 с.

Richards J. Remote Sensing. Digital Image Analysis. An Introduction. Second Edition. Springer-Verlag. Berlin Heidelberg New York London Paris Budapest. 1993. 340 p.

Бейтс Р., Мак-Доннелл М. Восстановление и реконструкция изображений. М.: Мир, 1989. 334 с.

Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложениях к интерферометрическим системам. СПб: ВНУ-Санкт-Петербург., 1998. 240 с.

Кашкин В.Б. Функциональные полиномы в задачах статистической радио техники. Новосибирск: Наука, 1982. 180 с.

4. РАСПОЗНАВАНИЕ ОБРАЗОВ

4.1. ОСНОВЫ ОБЩЕЙ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ

Под распознаванием образов (классификацией) обычно понимают процедуру, позволяющую вынести решение о принадлежности данного изображения или его фрагмента к одному из n классов, $n > 1$. Это решение выносится на основании наличия у изображения некоторых признаков того или иного класса. Результатом является выделение классифицированных объектов на изображении. Определение, что такое класс и каковы его признаки, зависит от задачи. На космическом изображении можно классифицировать области, покрытые снегом, точки зрения степени загрязнения их промышленными выбросами (классы — участки с разной степенью загрязнения). На этом же изображении можно выделить участки, покрытые лесом (классы — «лес» и «не-лес») и т.п. Распознавание образов относят к тематической обработке изображений, перед которой обычно выполняют предварительную обработку — восстанавливают и улучшают исходное изображение.

Задачи распознавания изображений с некоторой долей условности разделяют на задачи идентификации и задачи определения параметров положения. Например, опознание созвездия по изображению участка неба иного неба в системах астроориентации есть задача идентификации. Задача определения углов тангажа, рыскания и крена космического аппарата по положению опознанного созвездия на кадре есть задача определения параметров положения.

Иногда задачи идентификации и определения параметров положения объединяют в одну, называемую задачей распознавания. Задачи распознавания изображений решаются на основе общей теории распознавания образов.

Распознавание образов — раздел информатики, разрабатывающий принципы и методы классификации и идентификации предметов, явлений, процессов, сигналов, ситуаций, т.е. всех тех объектов, которые могут быть описаны конечным набором некоторых признаков или свойств, характеризующих объект.

В общей теории распознавания образов различают два типа задач:

- таксономии (обучение без учителя);
- собственно распознавания (обучение с учителем).

К задачам первого типа относят те, в которых предъявляемые объекты необходимо разделить на несколько групп (образов) только на

основе их описаний, причем число групп может быть известно или не известно.

К задачам второго типа относят те, в которых необходимо определить класс описания некоторого распознаваемого объекта. При решении этих задач предполагается, что число классов конечно и задано. Классы могут пересекаться.

Описание объекта представляет собой m -мерный вектор, где m — число признаков, используемых для характеристики объекта, причем q -й компонент этого вектора равен значению q -го признака. В описании объекта информация о значении некоторых признаков может отсутствовать, т.е. некоторые компоненты описания объекта могут быть не определены.

Наиболее общая математическая формулировка задачи распознавания выглядит следующим образом. Пусть для совокупности N объектов известны образы (классы), к которым они принадлежат. Эту совокупность называют обучающей последовательностью, а составляющие ее объекты — эталонами. Исходя из обучающей последовательности, необходимо определить класс, к которому принадлежит описание некоторого объекта, подвергаемого классификации и идентификации.

Формальное решение задачи распознавания принято записывать в виде так называемой решающей функции $i = d(\mathbf{x})$, где i — номер класса, к которому принадлежит описание распознаваемого объекта — вектор признаков \mathbf{x} .

Решающая функция строится на основе разбиения пространства признаков U на M непересекающихся классов $U_i, i = 1, \dots, M$:

$$\bigcup_{i=1}^M U_i = U, \quad U_i \cap U_j = \emptyset \text{ при } i \neq j.$$

В конечном итоге решающая функция d служит для определения номера i области U_i , к которой принадлежит аргумент \mathbf{x} . Критерием выбора решающей функции d может служить минимальная ошибка идентификации.

При числе классов более двух применяется метод последовательных дихотомий. Для наглядности вводят двоичное (бинарное) дерево распознавания, с помощью которого часто представляют множество данных при процедурах поиска, кодирования, распознавания и сортировки. Наиболее важным преимуществом использования двоичных деревьев в этих задачах является тот факт, что для идеально сбалансированного двоичного дерева число обращений к элементам исходного множества данных $\sim \log M$ вместо M при обычном переборе.

Двоичное дерево распознавания разбивает множество M возможных классов M_0 на два подмножества M_1 и M_2 , соответствующих крайним классам G_{i1} и G_{i2} множества M_0 :

$$\rho(G_{i1}, G_{i2}) = \max(\rho(G_i, G_j), M_0 = M_1 \cup M_2,$$

где ρ — функция расстояния на множестве M_0 .

Подмножество M_1 содержит векторы, более близкие к G_{i1} , чем к G_{i2} , подмножество M_2 , наоборот, содержит векторы, более близкие к G_{i2} , чем к G_{i1} . Подмножества M_1 и M_2 также разбиваются на два подмножества своих крайних элементов $M_1 = M_{11} \cup M_{12}$ и $M_2 = M_{21} \cup M_{22}$ и т.д.

Процесс поиска вектора признаков G_i , наиболее близкого к данному G_0 , можно оформить в виде последовательных дихотомий следующим образом. Находясь в корне дерева, производим сравнение расстояний $\rho(G_0, G_1)$ и $\rho(G_0, G_2)$, где G_1 и G_2 — векторы признаков, которыми помечены преемники корня. После этого переходим в ту вершину дерева, которая помечена ближайшим к G_0 вектором признаков (G_1 или G_2). Далее, находясь в этой вершине, продвигаемся по дуге, ведущей к ближайшему вектору, и так далее до тех пор, пока не достигнем листа дерева. После этого считаем, что наиболее близким вектором к G_0 из множества всех M эталонных векторов является вектор G_i , которым помечен достигнутый лист дерева распознавания.

4.1.1. Подходы к распознаванию изображений

В общей теории распознавания образов существует несколько различных подходов к распознаванию, отличающихся структурой описания классов признаков.

Детерминированный подход. Такой подход предполагает, что в любой точке пространства признаков, используемых для распознавания, с ненулевой априорной вероятностью могут появляться реализации только одного класса. При этом необходимо из всей совокупности признаков, характеризующих распознаваемые объекты, выделить систему детерминированных свойств, т.е. таких признаков, определение которых не связано с теми или иными принципиально случайными механизмами. Примерами таких свойств для изображений могут служить топологические инварианты, нормированные интегральные геометрические параметры, компоненты разложения изображения по какой-либо системе функций, моменты изображения и т.п.

Типичным примером систем распознавания изображений с использованием детерминированного подхода могут служить способы

распознавания типографских букв для ввода в ЭВМ текстов с помощью настольного сканера и их последующего анализа. Один из распространенных способов распознавания букв в этих системах основан на корреляционно-экстремальном алгоритме. Для этого в памяти ЭВМ должны храниться эталоны всех символов (букв, цифр, знаков препинания и др.), которые могут встретиться в обрабатываемом тексте. Каждый считываемый символ сравнивается со всеми эталонами, среди которых ищется максимально совпадающий с текущим считанным символом.

Непрерывно-групповой подход. Является развитием детерминированного подхода. Считается, что каждый рассматриваемый класс порождается в результате априори неизвестных топологических описываемых непрерывными группами преобразований некоторой эталонной реализации. В задаче обработки геометрически искаженных изображений, получаемых при дистанционных исследованиях Земли из космоса, эталонной реализацией служит географическая карта. Каждый класс в пространстве признаков является орбитой своей эталонной реализации, образуемой группой допустимых преобразований пространства признаков. При непрерывно-групповом подходе в качестве признаков выступают инварианты действующей в пространстве признаков группы. В рамках этого подхода можно непосредственно определять параметры положения эталонного объекта в предъявляемой реализации.

Из-за неопределенности положения один и тот же объект (например, озеро) может самым различным образом представляться на изображении. Для математического описания преобразований изображений используют математический аппарат теории непрерывных групп. Элементами группы являются различные преобразования изображений, а в качестве групповой операции выступает композиция этих преобразований.

При постановке задачи распознавания изображений в рамках непрерывно-группового подхода прежде всего необходимо установить, как из групп действует на распознаваемом изображении. Допустим, что это некоторая группа G . Рассмотрим некоторый эталонный объект S , наличие аналога которого на изображении мы предполагаем. На реальном изображении этот аналог будет выглядеть как gS , где g — априорно неизвестный элемент действующей на изображении группы G . При распознавании изображений в этих условиях обычно различают два типа (этапа) задач:

- идентификация объекта gS , выделенного на изображении, с эталоном S ;
- определение параметров положения объекта относительно датчика изображения, т.е. определение параметров (идентификация) преобразования g .

В данном случае существуют два основных подхода к решению задачи распознавания:

- с помощью нормализации;
- с помощью инвариантных признаков.

Нормализацией изображения, подвергнутого некоторому априорно неизвестному преобразованию g группы G , называют приведение изображения к так называемому нормализованному виду, т.е. к виду, совпадающему с соответствующим эталоном. Иначе говоря, нормализация — это устранение искажений, вызванных действием преобразования g . Затем нормализованное изображение отождествляют с одним из имеющихся эталонов. После этого, зная эталон S и его исходное изображение gS , вычисляют параметры преобразования g , т.е. решают задачу определения параметров положения.

При распознавании изображений с помощью нормализации важное место занимает понятие эквивалентности. Говорят, что изображения f_1 и f_2 эквивалентны относительно действия группы преобразований G , если найдется такое преобразование gG , при котором $gf_1 = f_2$. Множество всех изображений, эквивалентных (относительно действия группы преобразований G) заданному эталону S , называют орбитой S , а множество эквивалентных изображений — классом эквивалентности.

Простейшим примером использования нормализации может служить случай, когда $G = G_c$ — группа смещений. В данном случае классом эквивалентности эталона S является множество изображений, полученных из S путем различных сдвигов. Пусть на эталонах присутствует метка, указывающая на (условный) центр, при этом каждому классу эквивалентности соответствует один, и только один эталон. На исходном, предъявляемом для распознавания изображении $f(x, y)$ определяют центр (x_c, y_c) . Затем $f(x, y)$ нормализуют (центрируют), т.е. подвергают преобразованию g :

$$x' = x - (x_c - x_0), y' = y - (y_c - y_0),$$

где (x_0, y_0) — геометрический центр изображения. Полученное центрированное изображение $gf(x, y) = f(x', y')$ отождествляют с одним из имеющихся эталонов (которые также центрированы). Параметрами положения в данном примере являются координаты центра относительно геометрического центра изображения, т.е. $x_c - x_0$ и $y_c - y_0$.

Распознавание изображений с помощью инвариантных признаков предполагает переход из пространства изображений в пространство признаков. Выбор (синтез) признаков осуществляется так, чтобы эк-

вивалентным изображениям соответствовали одинаковые признаки и одинаковые векторы признаков. В условиях действия некоторой группы преобразований изображений признаки изображений должны быть инвариантны относительно преобразований данной действующей группы.

Общий теоретический прием синтеза системы инвариантных признаков и нормализаторов (преобразований нормализации) строится на основе теории дифференциальных уравнений в частных производных первого порядка.

Типичным примером задачи распознавания изображений с помощью инвариантных признаков может служить задача ориентации космического аппарата по изображениям участков звездного неба. При наличии в кадре k звезд с координатами (x_i, y_i) , $i = 1, \dots, k$, полная система элементарных инвариантных признаков для этого случая (в условиях действия группы $G_{\text{сн}}$ — смещения и поворота): \mathbf{r}_i^2 , $i = 1, \dots, k-1$; $\mathbf{r}_i \mathbf{r}_{i+1}$, $i = 1, \dots, k-2$, где \mathbf{r}_i — двумерный вектор, проведенный на изображении от k -й точки до i -й.

Структурно-лингвистический (синтаксический) подход. Основан на описании геометрической структуры класса с помощью специального формального языка, алфавит которого составляют так называемые непроизводственные элементы структуры и операции их композиции, задаваемые грамматикой этого языка. Такой подход предполагает, что структура и характер связей между элементами всех подлежащих распознаванию объектов подчиняются вполне определенным, заранее известным правилам. К настоящему времени разработано достаточно много приложений такого подхода, к которым относятся системы распознавания человеческой речи, автоматическая классификация некоторых микрообъектов (например, хромосом), распознавание геологических образований на космических изображениях Земли и др.

Все методы распознавания, основанные на детерминированном статистическом и непрерывно-групповом подходах, предполагают разделение объектов в пространстве признаков. При синтаксическом (иногда называемом грамматическим) подходе ищут и используют правила, которым подчиняется структура рассматриваемых образов. При этом процедуры классификации и идентификации состоят из двух этапов и являются достаточно простым и естественным следствием из предположения, что конструкция, или структура, всех подлежащих распознаванию объектов подчиняется вполне определенным, заранее известным правилам. На первом этапе разрабатываются те правила, которым подчиняются конструкции; иначе говоря, создается своего рода грамматика в том смысле, в котором этот термин используется в лингвистике.

Когда грамматика определена, наступает второй этап, состоящий в принятии решения о том, принадлежит ли рассматриваемый объект множеству всех объектов, порождаемых этой грамматикой.

В основе структурно-лингвистического подхода к распознаванию лежат основные понятия математической лингвистики, которая, в свою очередь, базируется на самых общих принципах построения различных конструкций в естественных языках.

Когда говорят о синтаксическом распознавании, то обычно имеют в виду применение этого термина к одномерным структурам. Термин «структурное распознавание» используют для двумерных (2D-данных) пиксельных данных, в частности изображений. Для трехмерных структур (3D-структур) чаще применяют термины «анализ сцен» и «распознавание сцен».

Структурный подход к представлению и описанию образов можно проиллюстрировать на примере. Пусть алфавит, т.е. множество элементарных примитивных объектов (примитивов), из которых могут состоять реальные объекты на изображении, содержит 4 элемента: $\Sigma = \{a, b, c, d\}$, где a — стрелка вверх \uparrow , b — стрелка вправо \rightarrow , c — стрелка вниз \downarrow , d — стрелка влево \leftarrow . Из этих примитивов можно составить различные фигуры, которым соответствуют слова и предложения с использованием алфавита. Слово $b^m c^n d^m a^n$ означает, что речь идет о прямоугольнике, горизонтальные стороны которого содержат m одинаковых отрезков, а вертикальные — n отрезков.

К линейным изображениям можно применить разнообразные процедуры кодирования контуров, в том числе так называемое цепочечное (цепное) кодирование по методу Фримена, когда алфавит образован достаточно большим числом элементов, подобных упомянутому, но назначенных под различными углами.

При структурном подходе под распознаванием понимается установление соответствия между предъявляемым объектом (точнее, его описанием) и одним из слов языка, порожденного используемой грамматикой.

Статистический подход. Предполагает, что в одной точке пространства признаков с отличной от нуля априорной вероятностью могут появляться реализации, относящиеся к различным классам. Это может быть связано с неизбежными случайными ошибками, присутствующими в детерминированных признаках, а также с использованием признаков, являющихся принципиально случайными величинами, т.е. вероятностных свойств.

Методы статистического распознавания образов наиболее часто применяют в задачах дистанционного зондирования в силу следующих причин:

— вследствие случайного характера протекания природных явлений результаты дистанционного зондирования подвержены случайным изменениям, маскирующим характерные различия между классами. Статистический анализ позволяет учесть эти изменения, потенциально уменьшить их отрицательное влияние на точность классификации;

— исследуемые классы могут перекрываться в пространстве измерений, т.е. некоторые измерения одного класса могут быть неотличимы от измерений других классов. В этих случаях методы статистического распознавания образов позволяют производить наиболее четкую классификацию.

Методы статистического распознавания образов обычно используют функции распределения вероятностей, связанные с классами образов. Однако зачастую эти функции неизвестны и должны оцениваться по множеству обучающих образов. В некоторых случаях форма функций распределения вероятностей считается известной (например нормальной) и по обучающим образам необходимо оценить только отдельные параметры, связанные с этими функциями (такие как математические ожидания, дисперсии, функции корреляции). Такой метод называется параметрическим. Если форма функций распределения вероятностей неизвестна заранее, метод является непараметрическим. Параметрические методы обычно легче реализуются, но требуют большего объема априорной информации или фундаментальных предположений относительно природы образов.

Непараметрические методы имеют большие потенциальные возможности для точной оценки функций распределения вероятностей и для точного распознавания, но это преимущество обычно требует сложных распознающих систем, большого числа обучающих образов и, главное, больших временных затрат.

Нейрокомпьютерный подход. Служит мощным инструментом для решения задач распознавания изображений. Основной структурной единицей нейрокомпьютера является формальный нейрон, суммирующий входные сигналы с определенными весовыми коэффициентами и выдающий результирующий сигнал на входы других нейронов и на выход нейронной сети. Обучение нейрокомпьютера сводится к подбору весовых коэффициентов для получения требуемого результата. Существует принципиальное отличие нейрокомпьютерного подхода от классических детерминированного и статистического. В данном случае интересна не конкретная конструкция устройства или конкретный алгоритм (они, как правило, остаются неизвестными), а положительный результат распознавания, который достигнут с помощью обучения и перестройки связей нейронов в процессе обучения.

4.2. СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ

Одним из самых распространенных методов выделения объектов на космических изображениях Земли является сегментация. Этот метод носит черты и детерминированного, и статистического подхода. Под сегментацией, в широком смысле, понимают преобразование полутоновых или цветных изображений в изображения, имеющие меньшее число тонов или цветов, чем исходные. В узком смысле сегментацией называют преобразование полутонового изображения в двухуровневое (бинарное), содержащее всего два уровня яркости — минимальный (обычно это 0) и максимальный (обычно 255). При этом объект и фон разделены, легко определить число объектов, характеристики их местоположения (координаты, поворот выделенной оси объекта относительно координатных осей и т.п.), геометрические характеристики (например, площадь каждого объекта, периметр, средний, минимальный, максимальный размеры) и, наконец, идентифицировать объект — указать, что это такое.

Целью сегментации является выделение областей (сегментов), однородных в каком-то определенном заданном смысле. Однородность является признаком принадлежности области к определенному классу.

Очень часто сегментация используется для выделения областей приблизительно одинакового тона и/или цвета, а также областей, однородных в смысле некоторого более сложного свойства (например, типа текстуры). Такие области принято называть *кластерами*.

Текстурой в теории обработки изображений называют структуру, которая характеризуется наличием повторяющегося «рисунка», состоящего из некоторых однородных участков приблизительно одинаковых размеров. Примером текстурного изображения являются фотоснимок кирпичной стены, аэрофотоснимок городских кварталов, космическое изображение участка летней тундры с многочисленными круглыми озерами.

Текстурной принято называть матрицу совместной встречаемости (вероятности) пар уровней серого тона у соседних пикселей. Пусть, например, значения элементов изображения f_{ij} представлены N градациями (обычно $N = 256$). Тогда текстурной матрицей T размером $N \times N$ считается та, (k, l) -й элемент которой есть вероятность (относительная частота) того, что соседние пары пикселей имеют значения f_k и f_l . Так как $I_{kl} = I_{lk}$, то текстурная матрица симметрична. В качестве характеристик текстуры могут рассматриваться различные скалярные характеристики текстурной матрицы.

Применяются три основных способа сегментации изображений: пороговая, путем наращивания областей, путем выделения границ.

4.2.1. Пороговая сегментация и кластерный анализ

Пороговая сегментация состоит в простом объединении близких по характеристикам областей изображения в небольшое число сегментов на основе априорно заданных порогов. Если яркость превышает порог, то элемент изображения относят к одному сегменту, если она ниже порога — то к другому. Этот самый простой способ основан на детерминированном подходе и требует минимальных вычислительных затрат.

Другой, более адекватный, способ выбора порогов заключается в том, что за пороги принимают границы мод гистограммы изображения. Рассмотрим более подробно этот способ выбора на примере бинаризации полутонового изображения, у которого гистограмма содержит две моды.

Если моды гистограмм не перекрываются или перекрываются слабо, то выбор порога разбиения изображения на две области U_1 и U_2 не представляет труда. Этот случай типичен для задачи выделения площадей, покрытых снегом и льдом на фоне леса и оттаявшей земли, по результатам дистанционных исследований. Гистограмма такого изображения имеет две моды: одна соответствует более темному фону, вторая — объектам с большей яркостью, т.е. снегу/льду, между модами существует резкая и протяженная зона минимума. Порог можно выбрать посередине зоны (рис. 4.1).

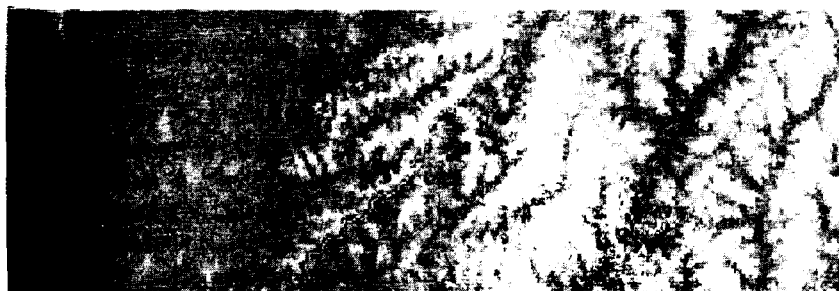
Сложнее провести сегментацию, когда гистограммы фона и объекта имеют перекрывающиеся «хвосты» и возможно, что при сегментации объект может быть отнесен к фону, а фон может быть принят за объект. Таким образом, может быть принято два решения: $d = 1$ или $d = 2$, соответствующие отнесению яркости текущего пиксела f к первой либо ко второй области. Пусть $d = 1$ отвечает $f < b$, $d = 2$ отвечает $f > b$, где b — порог.

Будем искать решающее правило d , минимизирующее функцию

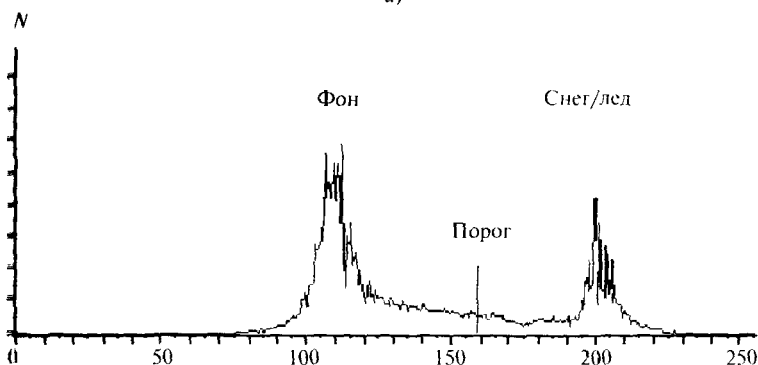
$$\langle c(i, d) \rangle = c_{11} p_{11} + c_{12} p_{12} + c_{21} p_{21} + c_{22} p_{22}, \quad (4.1)$$

где p_{ij} — вероятность принятия решения j , если на самом деле $f \in U_i$. Для учета возможных ошибок введем функцию потерь $c(i, d)$, представляющую собой количественную меру потерь, возникающих при принятии решения d , если на самом деле f принадлежит классу U_i . Функцию $c(i, d)$ иногда также называют функцией (или матрицей) штрафов (штрафной функцией) или риска.

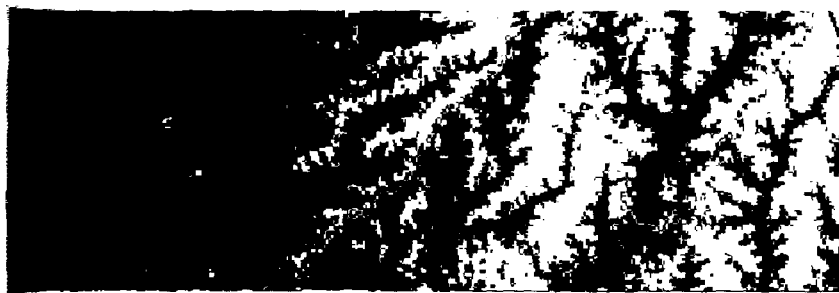
В случае двух классов функция потерь принимает 4 значения: c_{11} — когда принято решение, что f принадлежит классу U_1 , если f действительно



а)



б)



в)

Рис. 4.1. Исходное изображение (а), гистограмма уровней его яркости (б) и пороговое изображение (в)

но принадлежит этому классу; c_{22} — когда принято решение, что f принадлежит классу U_2 , если f действительно принадлежит этому классу; c_{12} — когда принято ошибочное решение отнесения первого класса ко второму; c_{21} — отнесению второго класса к первому.

Пусть p — априорная вероятность появления первого класса, тогда вероятность появления второго класса будет $1 - p$, $w_1(f)$ — плотность

вероятности распределения яркости пикселей первого класса, $w_2(f)$ – второго. Плотности вероятности $w_1(f)$ и $w_2(f)$ считаются известными или могут быть найдены из гистограммы. Средние суммарные потери записываются в виде

$$\begin{aligned} <c(i, d)> = c_{11} \int_0^b w_1(f) df + c_{12} p \int_b^\infty w_1(f) df + c_{21} (1-p) \int_0^b w_2(f) df + \\ + c_{22} (1-p) \int_b^\infty w_2(f) df. \end{aligned}$$

Величины c_{ij} , $i, j = 1, 2, \dots, p$, считаются известными. Неизвестной является граница b (порог) в пространстве признаков U , разделяющая классы U_1 и U_2 .

Оптимальное в смысле минимума суммарных средних потерь решающее правило определяется значением порога b , при котором $Q = <c(i, d)>$ имеет минимум. Это значение можно найти из условия $dQ/db = 0$, т.е.

$$\begin{aligned} dQ/db = (c_{11} - c_{12})p w_1(b) + (c_{11} + c_{12})p + (c_{21} - c_{22})(1-p)w_2(b) + \\ + (c_{21} + c_{22})(1-p) = 0. \end{aligned}$$

Отсюда получаем уравнение для определения порога:

$$\Lambda(b) = \Lambda_c = w_2(b)/w_1(b) = p(c_{11} - c_{12})/(1-p)(c_{22} - c_{21}).$$

Решив это уравнение, найдем оптимальное в смысле минимума средних потерь значение граничного порога b . Величина $\Lambda(f) = w_2(f)/w_1(f)$ называется отношением правдоподобия, а величина $\Lambda_c = w_2(b)/w_1(b)$ – критическим значением отношения правдоподобия

$$\Lambda_c = p(c_{11} - c_{12})/(1-p)(c_{22} - c_{21}).$$

Сравнивая отношение правдоподобия с Λ_c , находим решающее правило, соответствующее минимуму средних потерь:

$$\Lambda(f) \begin{cases} > \Lambda_c \Rightarrow d = 2, \\ < \Lambda_c \Rightarrow d = 1, \\ = \Lambda_c \Rightarrow \text{граница} \end{cases} \quad \text{или} \quad f \begin{cases} > b \Rightarrow d = 2, \\ < b \Rightarrow d = 1, \\ = b \Rightarrow \text{граница.} \end{cases}$$

В том случае, когда число классов больше двух, исходное множество U последовательным делением на два подмножества разбивают на некоторое число классов. Сначала выделяют два подмножества первого уровня U_1 и U_2 : $U_1 \cup U_2 = U$, $U_1 \cap U_2 = \emptyset$. Затем каждое подмножество U_1 , U_2 также разбивают на два подмножества второго уровня U_{11} , U_{12} , U_{21} и U_{22} :

$$U_{11} \cup U_{12} = U_1, U_{11} \cap U_{12} = \emptyset, U_{21} \cup U_{22} = U_2, U_{21} \cap U_{22} = \emptyset \text{ и т.д.}$$

При этом процесс принятия решения также сводится к последовательному отнесению f_k к одному из двух классов текущего уровня, т.е. реализуется метод последовательных дихотомий.

Кластерный анализ представляет собой обобщение пороговой сегментации. Кластерами называют элементы изображения (совокупности пикселей), в чем-то схожие между собой (по яркости, текстуре и т.п.). Пороговая сегментация в кластерном анализе основана на понятии расстояния, в выражение для которого может быть включена не только яркость, но и некоторые другие признаки. Если расстояние меньше порога, то элемент изображения относят к соответствующему кластеру. При использовании многоспектральных данных дистанционного зондирования в расстояние входят значения яркости (i, j) -го пикселя изображения в различных каналах. Совокупность этих значений можно записать в виде вектора f_{ij} . Кластеры формируются так, чтобы в каждом из них расстояние между отдельными пикселями было минимальным, а расстояния между пикселями, относящимися к различным кластерам, были как можно больше. Наиболее распространенной мерой подобия (сходства) является евклидово расстояние между векторами f_{ij} и f_{mn} . Если $\{f_{kij}\}$, $\{f_{kmn}\}$ — компоненты этих векторов, k — номер спектрального канала, то расстояние

$$r_{ij} = \sqrt{\sum_k (f_{kij} - f_{kmn})^2}.$$

В качестве меры сходства можно использовать также косинус угла между векторами

$$\cos \varphi = \frac{(f_{ij}, f_{mn})}{\|f_{ij}\| \|f_{mn}\|},$$

представляющий собой отношение скалярного произведения векторов к произведению их норм и максимальный при близости направлений векторов.

Процедура кластеризации может основываться на оптимизации какого-нибудь показателя качества, например на критерии минимума суммы квадратов ошибки:

$$\varepsilon = \sum_{k=1}^K \sum_{f \in S_k} \|f_k - \mu_k\|^2,$$

где K — число кластеров, S_k — множество объектов (пикселов), относящихся к k -му кластеру, μ_k — вектор средних значений для класса k .

В алгоритме ISODATA вначале достаточно произвольно выбираются K векторов в качестве различных стартовых точек. Затем производится кластеризация по минимуму расстояния от этих центров с использованием евклидовой метрики. После этого вычисляются векторы средних значений μ_k и находится средний квадрат ошибки ε . В следующем цикле выполняется повторная кластеризация по минимуму расстояния от векторов средних значений μ_k . При этом число кластеров может измениться, изменятся и векторы средних значений, и значение ε . При следующей итерации производится кластеризация по минимуму расстояния от новых векторов средних значений, вновь вычисляется

ε . Процесс продолжается до тех пор, пока величина ε не перестанет заметно уменьшаться.

Алгоритм ISODATA предусматривает управление процессом кластеризации: можно изменять число кластеров, число пар кластеров, которые можно объединить; допустимое число циклов итерации и т.п.

На рис. 4.2 приведен фрагмент изображения

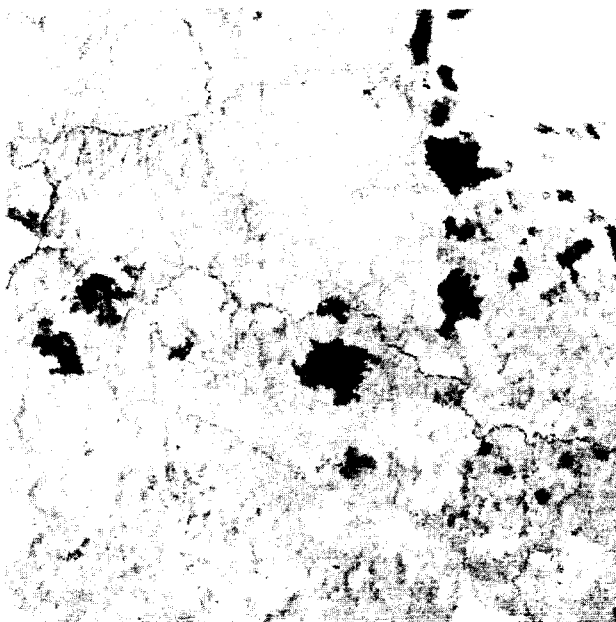
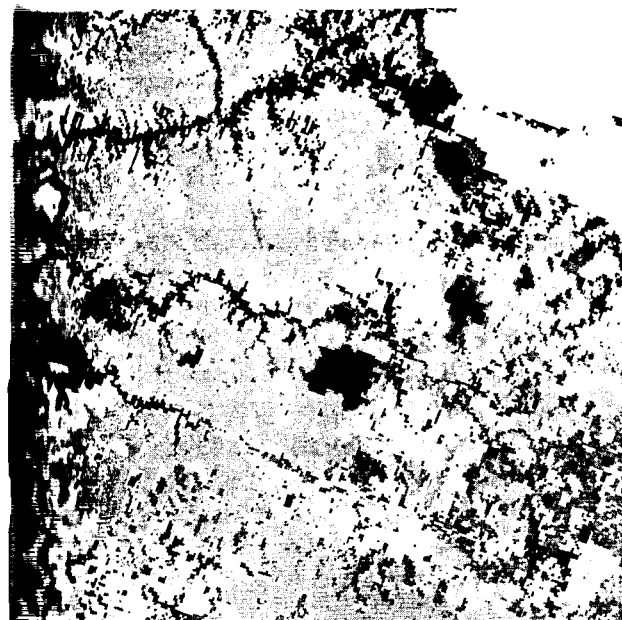
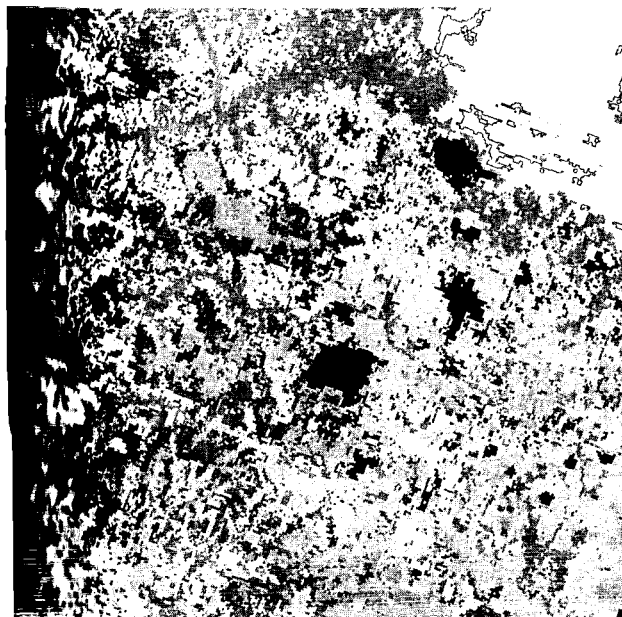


Рис. 4.2. Фрагмент изображения территории Эвенкии

1 2 3 4 5 6 7 8 9 10 11 12



а)



б)

Рис. 4.3. Изображение на рис. 4.2 после кластерного анализа с использованием алгоритма ISODATA без обучающей выборки (а) и с обучающей выборкой (б)

ния территории Эвенкии, полученный с помощью спутника NOAA-14 в 1-м канале сканера AVHRR. Для распознавания используется изображение в 1, 2, 5-м каналах. Изучаемый участок содержит пожарища в возрасте 1–3 года. На изображениях, полученных из космоса, легко определить свежевыгоревшие участки, но заросшие травой и кустарником гари прошлых лет относятся к числу достаточно сложных для распознавания объектов. Результат применения алгоритма ISODATA без использования обучающей выборки можно видеть на рис. 4.3, а.

Центры кластеров на первом шаге выбирались случайным образом. Были заданы параметры: число классов — 12, число итераций q — не более 30, процесс классификации прекращался, когда отношение $\delta = (\epsilon_q - \epsilon_{q-1})/\epsilon_q$ (порог сходимости) достигало 0,99. Здесь через ϵ_q и ϵ_{q-1} обозначены средние квадраты ошибки на q -м и $(q-1)$ -м шаге. Алгоритм отнес водную поверхность и тени от облаков к пожарищам.

На рис. 4.3, б после применения алгоритма ISODATA с обучающей выборкой видны основные присутствующие на изображении классы: 1–2-й — пожарища; 3–8-й — различные по отражательной способности типы леса; 9–10-й — облачность; 11-й — тень от облаков; 12-й — водная поверхность. Количество итераций и порог сходимости δ задавались так же, как и выше. Качество распознавания лучше, чем в предыдущем случае, все пожарища распознаны правильно.

4.2.2. Сегментация путем наращивания областей

При такой сегментации выделяются однородные области. Вначале рассмотрим этот способ, используя критерий однородности по значению яркости (вектора яркости).

Алгоритм предусматривает выбор стартового пиксела и проверку близости значений смежных с ним пикселей, например по евклидову расстоянию. Если значения яркости текущего и смежного с ним пикселей оказываются близкими, то их зачисляют в одну область. Таким образом, область формируется в результате сращивания отдельных пикселей. На определенном этапе (зависящем от модификации алгоритма) область проверяется на однородность. Если результат проверки оказывается отрицательным, то область разбивается на более мелкие участки. Процесс продолжается до тех пор, пока все выделенные области не выдержат проверки на однородность. Возможны реализации алгоритма, предусматривающие формирование областей сращиванием как отдельных пикселей, так и небольших областей.

Общая схема проверки области на однородность состоит в следующем. Пусть $F(R)$ — заданная мера однородности области R . Если

$R_{12} = R_1 \cap R_2$, то критерий однородности можно задать, потребовав, чтобы выполнялось условие $F(R_{12}) \leq \varepsilon$, где ε — заданный порог.

Различные алгоритмы сегментации прежде всего классифицируются по виду меры $F(R)$. В самых простых случаях в качестве F могут использоваться значения размаха

$$F(R) = \max_{(i,j) \in R} f_{ij} - \min_{(k,m) \in R} f_{km}$$

либо выборочные дисперсии

$$F(R) = \frac{1}{n} \sum_{(i,j) \in R} \left(f_{ij} - \frac{1}{n} \sum_{(k,m) \in R} f_{km} \right)^2,$$

где n — площадь (число пикселей) области R .

При обработке многоспектральных данных дистанционного зондирования в качестве $F(R)$ выбирают расстояние между пикселями в пространстве яркостей в спектральных каналах по метрике Евклида:

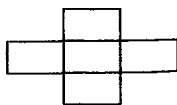
$$F(R) = [A_1(f_{1ij} - f_{1km})^2 + A_2(f_{2ij} - f_{2km})^2 + A_3(f_{3ij} - f_{3km})^2 + \dots]^{1/2}.$$

Здесь f_{1ij} — яркость (i, j) -го пикселя в 1-м канале, f_{2ij} — яркость во 2-м канале и т.д.; A_k — весовые коэффициенты, учитывающие зависимость яркостей объектов в каналах.

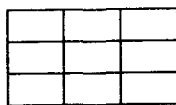
В более сложных вариантах метода сегментации путем наращивания областей могут быть использованы скалярные характеристики текстовой матрицы, например ее след.

Таким образом, при сегментации путем наращивания областей учитывается структура области, ее связность, что бывает важно при обработке данных дистанционного зондирования. Нередко этот метод дает лучшие результаты, чем другие, не учитывающие связность и рассчитанные на индивидуальное отнесение каждого пикселя к тому или иному классу.

Дальнейшая классификация алгоритмов основана на способе наращивания области. При использовании квадратной или прямоугольной сетки применяют два вида связности: 4- и 8-связность —



4-связность



8-связность

Приведем примеры использования алгоритмов наращивания областей.

Алгоритм построения связной области по строкам (или по столбцам) применяется только для выпуклых областей. Построение начинается с произвольно выбранной точки изображения.

Для односвязной (имеющей один контур) невыпуклой области больше всего подходит алгоритм наращивания контура, напоминающий правило одной руки при обходе лабиринта.

Для q -связной (имеющей q взаимно непересекающихся контуров) области также лучше всего использовать алгоритм наращивания принадлежащих q контуров.

В литературе отмечается, что если априори известно число областей сегментации и местоположение стартовых пикселей (которые должны находиться на достаточном расстоянии от границ областей), а также использован простой в вычислительном отношении критерий однородности сегментов, то метод наращивания областей прост в алгоритмическом отношении и обеспечивает хорошее качество сегментации. Это связано прежде всего с тем, что к определенному классу просто относят каждый пиксел в отдельности, как в методах, рассмотренных в п. 4.5.2–4.5.3, а учитывают окружение.

Результат применения алгоритма наращивания областей (исходное изображение на рис. 4.2) можно видеть на рис. 4.4. Область нара

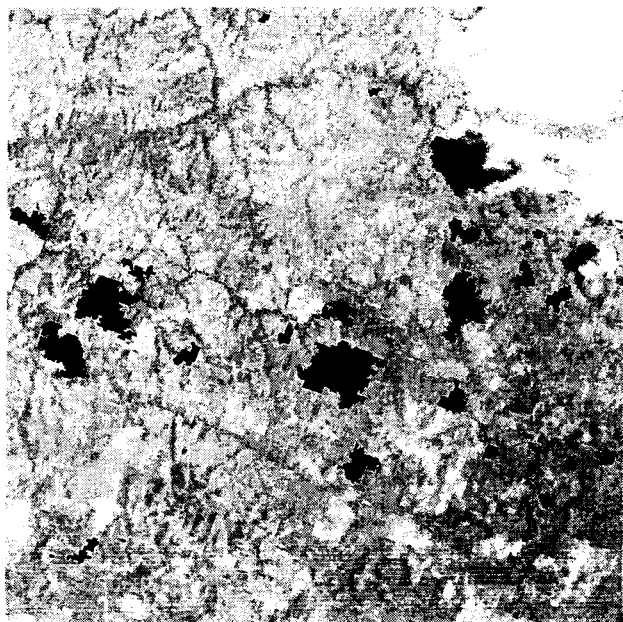


Рис. 4.4. Изображение после сегментации путем наращивания областей

инируется, начиная со стартовой точки, выбираемой оператором для каждой области вручную. Точки добавляются из восьми- или четырех-очной окрестности при условии, что евклидова дистанция между точкой и средним значением области лежит в заданных пределах. Евклидова дистанция $F(R)$ задавалась индивидуально для каждой области и лежала в пределах 2,2–5,6 (при квантовании яркости на 256 уровней), использовалась четырехсвязная окрестность. По сравнению с другими рассмотренными методами наращивание областей обеспечивает самое высокое качество распознавания.

3.3. Сегментация путем выделения границ

Сегментация путем выделения границ предусматривает использование оператора градиента. После этого для установления факта, что граница действительно обнаружена, применяется процедура разделения по порогу. Затем пикселы, идентифицированные как граничные, соединяются в замкнутые кривые, окружающие соответствующие области.

В этом методе, как и в других методах сегментации, существен критерий однородности области, по которому и вычисляются значения градиента. Прямые методы сегментации путем выделения границ предусматривают применение к исходному изображению градиентных фильтров (Роберта, Собела, Лапласа и др.).

Построение границ сегментов на изображении градиента выступает в качестве самостоятельной задачи, довольно сложной и решаемой лишь в самых простейших случаях. Например, можно выделять локальные максимумы градиента всех строк и столбцов изображения.

Существует разновидность метода сегментации путем выделения границ с использованием обратногоградиентных фильтров. В данном случае цель применения фильтров — повышение однородности сегментов изображения и резкости границ сегментов.

Как упоминалось в гл. 3, обратногоградиентная фильтрация заключается во взвешенном усреднении значений текущей окрестности, при котором весовые коэффициенты обратно пропорциональны модулю градиента изображения в текущей точке. В общем случае закон соответствия весовых коэффициентов и модуля градиента не обязательно обратно пропорциональный, важно лишь, что чем больше модуль градиента, тем меньше должен быть соответствующий весовой коэффициент, а зависимость весового коэффициента a от модуля градиента G должна быть невозрастающей.

Одним из преимуществ рассматриваемого метода является возможность построения границ с любой заданной точностью, которая может превосходить минимальное расстояние между соседними пиксела-

ми. Для этого можно интерполировать исходное изображение либо изображение градиента на более густую сетку.

Для границ простой формы (прямолинейных, дуг окружностей и т.п.) можно использовать аппроксимацию исходного изображения градиента какой-либо параметрической функцией. Примером такой задачи является ориентация искусственного спутника Земли по кадру изображения части края земного диска (поиск так называемой инфракрасной вертикали).

Сегментация путем выделения границ изображения на рис. 4.2 показала, что метод достаточно хорошо работает только при большой протяженности границы.

4.3. ОБНАРУЖЕНИЕ ОБЪЕКТОВ ИЗВЕСТНОЙ ФОРМЫ НА ИЗОБРАЖЕНИИ

Обнаружение (выделение) объектов известной формы на изображении можно считать одной из задач сегментации. При этом целью обработки является изображение, состоящее из сегментов $n+1$ классов – объектов n классов, выделенных на изображении, и фона.

Кроме полученного $(n + 1)$ -уровневого изображения, результатом выделения объектов может служить также вектор параметров — число выделенных объектов, их координаты и другие характеристики.

Вместе с тем существуют подходы, специфичные именно для задач выделения объектов на изображении, например звезд на кадре участка звездного неба, некоторого объекта (известного озера, острова и т.п.) на космическом изображении Земли. Предположим, что в окрестности D_{kl} текущего (i, j) -го пиксела присутствует одиночный выделяемый объект (импульс сложной формы) s_{ijkl} . Конечно, окрестность D_{kl} должна быть не меньше размера объекта s_{ijkl} , форма которого, зависящая в общем случае от положения на изображении, считается известной. Идеальную форму объекта s_{ijkl} называют эталонным импульсом. Не нарушая общности, амплитуду (максимальное значение) эталонного импульса s_{ijkl} можно считать равной единице. Объект наблюдается совместно с фоном $p_{ijk}(\lambda)$, который зависит от λ — вектора случайных параметров. Задача обработки состоит том, чтобы оптимальным образом определить, присутствует или нет объект в окрестности D_{kl} .

Самая простая модель фона $p_{ijkl} = \text{const}$ в окрестности D_{kl} . Более реалистична модель квазидетерминированного фона, примером которого может служить неравномерно освещенная поверхность. Характерный размер объекта, который предстоит выделить на изображении

Фон мал по сравнению с характерным размером неоднородности фона. В этом случае в пределах текущей окрестности D_{kl} точки (k, l) фон можно считать линейно изменяющимся, его значение в окрестности

$$p_{ijkl} = a_{kl} + b_{kl} \cdot i + c_{kl} \cdot j,$$

где a_{kl} , b_{kl} и c_{kl} — некоторые константы, значения которых априорно неизвестны. Координаты i и j по строкам и столбцам здесь выступают как детерминированные величины. Константы a_{kl} , b_{kl} и c_{kl} оцениваются по исходному изображению, в данном случае их следует считать случайными величинами. Коэффициент a_{kl} отвечает за постоянный фон, b_{kl} и c_{kl} — за линейное изменение яркости фона вдоль строк и столбцов. В этом примере фон есть детерминированная функция детерминированных и случайных величин, т.е. его можно считать квазидетерминированным случайным полем.

Кроме того, на наблюдаемый пиксел f_{ij} накладывается аддитивный шум ϵ_{ijkl} , носящий случайный характер и имеющий нормальный закон распределения. К ϵ_{ijkl} можно также отнести ошибки измерений. Таким образом, наблюдаемое изображение есть

$$f_{ij} + \epsilon_{ijkl} = p_{ijkl}(\lambda) + h_{kl} \cdot s_{ijkl} = a_{kl} + b_{kl} \cdot i + c_{kl} \cdot j + h_{kl} \cdot s_{ijkl}, \quad (i, j) \in D_{kl}, \quad (4.2)$$

где h_{kl} — амплитуда импульса s_{ijkl} на изображении. Задачу обнаружения (выделения) объекта можно свести к определению значения h_{kl} , которое затем сравнивают с заранее определенным порогом. Если h_{kl} превышает порог, принимается решение о наличии объекта s_{ijkl} в текущей окрестности $(i, j) \in D_{kl}$. Среднеквадратическая ошибка аппроксимации

$$\Phi_{kl} = \frac{1}{N} \sum_{(i,j) \in D_{kl}} \epsilon_{ijkl}^2 = \frac{1}{N} \sum_{(i,j) \in D_{kl}} (f_{ij} - p_{ijkl}(\lambda) - h_{kl} \cdot s_{ijkl})^2,$$

где N — число точек окрестности D_{kl} . Условие минимума Φ_{kl} записывается как

$$\partial \Phi_{kl} / \partial \lambda_1 = \dots = \partial \Phi_{kl} / \partial \lambda_n = \partial \Phi_{kl} / \partial \lambda,$$

где n — число параметров фона.

Такой общий подход называют оптимальным обнаружением, а формальный оператор $h_{kl} = \mathbf{F}(\{f_{ij}; (i, j) \in D_{kl}\})$ — оптимальным обнаружите-

лем (оптимальность понимается в смысле минимума среднеквадратической ошибки аппроксимации). Рассмотрим частные случаи.

1. Пренебрежимо малый фон. В этом случае система уравнений (4.2) имеет вид

$$h_{kl} \cdot s_{ijkl} = f_{ij} + \varepsilon_{ijkl}, \quad (i, j) \in D_{kl}.$$

Решение этой системы

$$h_{kl} = \{sf\} / \{s^2\}.$$

Здесь для суммирования используется обозначение Гаусса, применяемое в методе наименьших квадратов:

$$[xy] = \sum_{(i,j) \in D_{kl}} x_{ijkl} y_{ijkl}.$$

2. Постоянный в пределах текущей окрестности фон. Из исходной системы условных уравнений $a_{kl} + h_{kl} \cdot s_{ijkl} = f_{ij} + \varepsilon_{ijkl}, (i, j) \in D_{kl}$, получается система нормальных уравнений метода наименьших квадратов

$$\begin{bmatrix} [1] & [s] \\ [s] & [s^2] \end{bmatrix} \begin{bmatrix} [a_{kl}] \\ [h_{kl}] \end{bmatrix} = \begin{bmatrix} [f] \\ [fs] \end{bmatrix}.$$

3. Линейно изменяющийся фон. Систему условных уравнений метода наименьших квадратов (4.2) преобразуем в систему нормальных уравнений:

$$\begin{bmatrix} [1] & [i] & [j] & [s] \\ [i] & [i^2] & [ij] & [is] \\ [j] & [ij] & [j^2] & [js] \\ [s] & [is] & [js] & [s^2] \end{bmatrix} \begin{bmatrix} [a_{kl}] \\ [b_{kl}] \\ [c_{kl}] \\ [h_{kl}] \end{bmatrix} = \begin{bmatrix} [f] \\ [if] \\ [jf] \\ [sf] \end{bmatrix},$$

определитель $d = [s]A + [is]B + [js]C + [s^2]D$ которой представляет собой разложение по 4-му столбцу, где A, B, C и D — алгебраические дополнения соответствующих элементов 4-го столбца. Амплитуду h_{kl} также можно записать в виде разложения по 4-му столбцу:

$$h_{kl} = (1/d) ([f]A + [if]B + [jf]C + [sf]D).$$

Если форма эталонного импульса s_{ijkl} не зависит от положения на изображении, т.е. определяется только i и j , то преобразование $f_{ij} \rightarrow h_{kl}$ линейное (линейная фильтрация). Этот фильтр вместе с пороговым устройством носит название оптимального линейного обнаружителя.

В случае постоянного фона амплитуда

$$h_{kl} = \frac{[1][fs] - [f][s]}{[1][s^2] - [s]^2}. \quad (4.3)$$

Знаменатель этого выражения — константа, а числитель пропорционален коэффициенту корреляции эталонного импульса s_{ij} и исходного изображения f_{ij} . Отсюда следует, что в качестве меры наличия объекта в текущей окрестности можно выбрать коэффициент корреляции r_{sf} между s и f . При этом осуществляют поиск локальных максимумов $(r_{sf})_{kl}$ и, сравнивая их значения с заранее выбранным порогом, принимают решение о наличии или отсутствии объекта в текущей окрестности изображения. Данный подход, называемый *корреляционно-экстремальным*, может применяться при автоматическом совмещении спутникового изображения с картой. Пример его использования см. в п. 5.9.5.

4.4. ПРИМЕНЕНИЕ ТЕОРЕМЫ БАЙЕСА. РЕШАЮЩЕЕ ПРАВИЛО

При статистическом подходе к распознаванию часто используется теорема Байеса — одна из фундаментальных в теории вероятностей. Рассматриваются *случайные события*, т.е. такие, для каждого из которых невозможно заранее предсказать точный результат. Однако при большом числе реализаций эти события можно характеризовать средними результатами, стабильными и воспроизводимыми. Если могут произойти два события A и B , то можно говорить о трех различных вероятностях:

- событие A происходит с вероятностью $P(A)$;
- событие B происходит с вероятностью $P(B)$;
- события A и B происходят одновременно с вероятностью $P(A \cap B)$, называемой вероятностью совместного события или совместной вероятностью.

Пусть A_i ($i = 1, 2, \dots, n$) — случайное событие, вероятное при условии, что произошло случайное событие B . Эту условную вероятность называют *апостериорной* и вычисляют по теореме Байеса:

$$P(A_i / B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B / A_i)}{\sum_i P(A_i)P(B / A_i)}, \quad (4.4)$$

поскольку $P(B) = \sum_i P(B/A_i)P(A_i)$. Здесь $P(A_i)$ — *априорная вероятность* события A_i .

Используя апостериорные вероятности, можно разработать различные методы автоматической классификации. Пусть \mathbf{X}_k — вектор измерений, представляющий k -й класс. Априорная вероятность того, что эта реализация относится к классу с номером k , есть $P(\mathbf{X}_k)$; она считается заданной самой постановкой задачи. Например, для лесных территорий бывает известен процент площади, занятой лиственными и хвойными породами.

Необходимо отнести неизвестный наблюдаемый объект X к одному из известных классов C_k , с минимальной ошибкой. Для этого вычисляют m измерений в соответствии с выбранными признаками. Результатом является вектор измерений \mathbf{X}_m , для которого можно найти условную вероятность (или ее плотность):

$$\begin{array}{c} P(A_i/B) \\ \uparrow \uparrow \\ \text{Вероятное Известное} \end{array}$$

Решение об отнесении неизвестного объекта к классу с номером k можно считать оправданным, если выполняется условие:

$$P(C_k/\mathbf{X}_m) \geq P(C_j/\mathbf{X}_m), j \neq k.$$

Эти вероятности могут быть вычислены согласно теореме Байеса с использованием условных вероятностей $P(\mathbf{X}_m/C_k)$:

$$P(C_k/\mathbf{X}_m) = \frac{P(C_k)P(\mathbf{X}_m/C_k)}{P(\mathbf{X}_m)}, \quad P(C_j/\mathbf{X}_m) = \frac{P(C_j)P(\mathbf{X}_m/C_j)}{P(\mathbf{X}_m)}, \quad (4.5)$$

откуда следует решающее правило — необходимо определить, какой из апостериорных вероятностей больше:

$$P(C_k) P(\mathbf{X}_m/C_k) \text{ или } P(C_j) P(\mathbf{X}_m/C_j). \quad (4.6)$$

Процесс классификации связан с некоторым риском, так что возможны ошибки. Для их учета, как и в п. 4.3.1, введем *функцию потерь* — цену принятия правильного или ошибочного решения, называемую также *функцией штрафов* (штрафной функцией) или *функцией риска*

Пусть d — правило, по которому производится классификация. Решение d_i означает, что объект X_m отнесен к классу C_i . Пусть $c(d_i/C_j)$ — стоимость принятия решения d_i , если известно, что объект принадлежит классу C_j . Средний риск принятия решения d_i в случае, когда находится объект X_m ,

$$R(d_i/X_m) = \sum_{j=1} c(d_i/C_j)P(C_j/X_m).$$

Средний суммарный риск, связанный с решающим правилом d для реализаций X_m ,

$$R(d) = \int_{X_m} R(d/X_m)p(X_m)dX_m, \quad p(X_m) = dP(X_m)/dX_m.$$

Оптимизация процесса классификации сводится к отысканию такого решающего правила d^* , которое минимизировало бы средний суммарный риск:

$$R^* = R(d^*) = \min_d R(d).$$

Пример использования критерия минимума среднего суммарного риска при пороговой сегментации приведен в п. 4.4.1. Кроме этого критерия, используются и другие. Например, из всех вариантов выбирается такой, который отвечает минимальному среднему риску; может потребоваться получение максимальной вероятности правильной классификации или минимума средней ошибки классификации и т.д.

Рассмотрим классификацию по критерию минимального среднего риска. Пусть рассматриваемый объект может быть отнесен к одному из n классов. Обозначим: $c_{11} = c(d_1/C_1)$ — функция риска (цена) принятия решения, что X_m принадлежит классу C_1 , если X_m действительно принадлежит этому классу; $c_{22} = c(d_2/C_2)$ — соответствует принятию решения, что X_m принадлежит классу C_2 , если X_m действительно принадлежит этому классу; $c_{12} = c(d_1/C_2)$ — ошибочному решению отнесения первого класса ко второму; $c_{21} = c(d_2/C_1)$ — отнесению второго класса к первому. Средние цены принятия каждого решения:

$$R_1(X) = c_{11}P(C_1/X_m) + c_{12}P(C_2/X_m),$$

$$R_2(X) = c_{21}P(C_1/X_m) + c_{22}P(C_2/X_m).$$

Пусть риск принятия первого решения меньше риска принятия второго: $R_1(X) < R_2(X)$, т.е.

$$c_{11}P(C_1/X_m) + c_{12}P(C_2/X_m) < c_{21}P(C_1/X_m) + c_{22}P(C_2/X_m).$$

Положим $c_{11} < c_{12}$ и $c_{22} < c_{21}$, т.е. что риск принятия верного решения меньше риска допустить ошибку. Тогда

$$(c_{11} - c_{21}) P(C_1/X_m) < (c_{22} - c_{12}) P(C_2/X_m),$$

$$(c_{21} - c_{11}) P(C_1/X_m) > (c_{12} - c_{22}) P(C_2/X_m).$$

Применяя формулу Байеса, получаем решающее правило в виде неравенства

$$\frac{P(X_m/C_1)}{P(X_m/C_2)} > \frac{(c_{12} - c_{22})P(C_2)}{(c_{21} - c_{11})P(C_1)}. \quad (4.7)$$

В левой части неравенства стоит дробь, которую называют *отношением правдоподобия*. Решение о том, что наблюдается объект первого класса, принимается тогда, когда отношение правдоподобия превышает выражение в правой части. Правая часть содержит только известные члены и представляет собой значение порога, не зависящее от наблюдаемой величины.

Пусть цена имеет вид $c_{ij} = 1 - \delta_{ij}$, где $\delta_{ij} = 1$, $i = j$; $\delta_{ij} = 0$, $i \neq j$. В этом случае вместо (4.7) получаем

$$\frac{P(X_m/C_1)}{P(X_m/C_2)} > \frac{P(C_2)}{P(C_1)}. \quad (4.8)$$

Подсчитаем вероятность ошибки, которая может возникнуть при использовании этой процедуры. Пусть $P(C_1) = P(C_2)$. На рис. 4.5 показаны две кривые плотности вероятности, представляющие зависимость $w(X_m/C_1)$ и $w(X_m/C_2)$ от X_m . На прямой $A-A$ выполняется неравенство (4.8), поэтому можно считать, что X_m принадлежит к классу C_1 .

Пусть линии раздела $D-D$ соответствует граничное значение абсциссы X^* . Если перемещать линию $A-A$, разделяющую два решения, вдоль оси X_m до совпадения с линией $D-D$, то она достигнет точки X^* , в которой имеет место равенство $P(X^*/C_1) = P(X^*/C_2)$. Любая точка, для которой $X_m < X^*$, считается принадлежащей к классу C_1 , в то время как

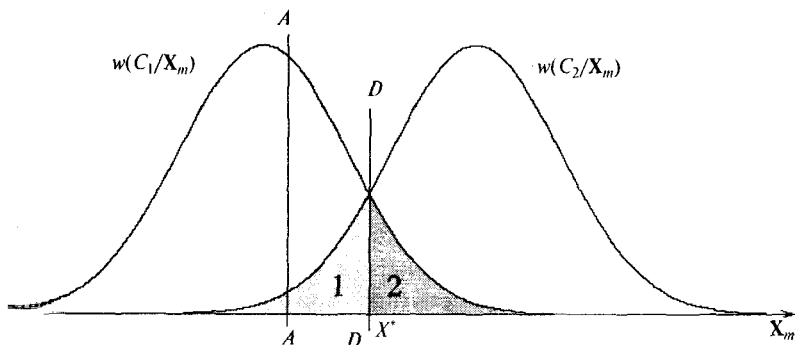


рис. 4.5. Условные плотности вероятности (разделение на два класса)

Для точек, для которых $X_m > X^*$, относятся к классу C_2 . Однако вероятность того, что в первом случае точка может принадлежать к классу C_2 , отлична от нуля (область 1), как и то, что во втором случае точка X^* принадлежит к классу C_1 (область 2). Для класса C_1 область 1 является зоной ложной тревоги (л.т), а область 2 — зоной пропуска обнаружения (п.о). Вероятности ложной тревоги $P_{л.т}$ и пропуска обнаружения $P_{п.о}$ равны соответственно площадям под областями 1 и 2 на рис. 4.6. Они определяются соответственно выражениями

$$P_{п.о} = \int_{-\infty}^{X^*} w(X_m / C_2) dX_m,$$

$$P_{л.т} = \int_{X^*}^{\infty} w(X_m / C_1) dX_m.$$

Суммарная ошибка классификации равна сумме двух вероятностей:
 $P_{пш} = P_{л.т} + P_{п.о}.$

4.5. ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ КЛАССИФИКАЦИИ

4.1. Нормальный закон распределения

Задачах классификации

Изображение земной поверхности, полученное дистанционными методами, представляет собой некоторое случайное поле. Яркость каждого пикселя изображения складывается из яркостей природных объ-

ектов, попавших в мгновенное поле зрения сканера. Объекты по-разному, случайным образом, отражают и рассеивают солнечное излучение, так что яркость каждого пиксела есть сумма случайных величин. Из теории вероятностей известно, что сумма независимых одинаково распределенных случайных величин имеет нормальный (гауссовский) закон распределения при условии, что слагаемые вносят равномерно малый вклад в сумму (центральная предельная теорема). Существуют обобщения центральной предельной теоремы на сумму зависимых случайных величин, на сумму случайных величин с различными законами распределения и т.п. Таким образом, есть основания полагать, что для природных объектов яркость пикселей подчиняется нормальному закону распределения.

Названная причина не единственная, по которой в задачах параметрического распознавания изображений широко используется гипотеза о нормальном законе распределения яркости пикселей. При распознавании монохроматических изображений для распределения яркости почти всегда удается подобрать один из известных одномерных законов (нормальный, биномиальный, экспоненциальный, пуассоновский и др.). Иное дело изображения, получаемые с помощью сканеров, имеющих $n > 1$ спектральных каналов. В этом случае для описания статистических характеристик требуется n -мерное совместное распределение яркости в каналах. Как указывают многие авторы, в настоящее время уровень знаний таков, что пока точному многомерному анализу, за редкими исключениями, поддаются лишь те задачи, где рассматривается нормальный многомерный закон распределения.

Существуют и чисто технические причины. Для записи в памяти ЭВМ гистограммы, представляющей функции распределения вероятностей яркости пикселей изображения, может потребоваться большой объем памяти. Число ячеек, необходимых для записи n -мерной гистограммы, в которой каждое значение яркости может принимать p уровней, равно p^n . Один из способов разрешения этой трудности — предположить, что каждая гистограмма, или функция распределения вероятностей, может быть адекватно аппроксимирована кривой, имеющей простую функциональную форму и зависящей от небольшого числа параметров. В частности, можно считать, что функция распределения вероятностей для любого интересующего класса может быть представлена нормальной функцией.

Однако нормальная плотность вероятности предполагает, что значения яркости могут быть отрицательными, хотя яркость всегда положительна; значения яркости в цифровом формате всегда ограничены сверху. Нормальная плотность вероятности описывается одномодальной (имеющей один максимум) кривой. В то же время гистограмма

представляющие функции распределения вероятностей яркости пикселей изображения, могут быть многомодальными. Поэтому всегда следует помнить, что нормальная кривая дает приближенное описание плотности вероятности яркости изображения.

Для одномерного случая нормальная функция плотности вероятности яркости пикселей f_{ij} для некоторого класса k имеет вид

$$w_1(f_{ij}) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp[-(f_{ij} - \mu_k)^2 / 2\sigma_k^2],$$

где (i, j) — координаты пикселя; μ_k — математическое ожидание (среднее значение) яркости для класса k ; $\sigma_k^2 = \langle (f - \mu)^2 \rangle$ — дисперсия яркости пикселей для класса k . Ниже используется также обозначение $\sigma_{1k} = \sigma_k^2$.

На практике μ_k и σ_k^2 неизвестны и должны оцениваться по обучающим выборкам. Несмещенные оценки для μ_k и σ_k^2 определяются выражениями

$$\bar{\mu}_k = \frac{1}{Q} \sum_{i=1}^Q f_{m_i}, \quad (4.9)$$

$$\sigma_k^2 = \bar{R}_{1k} = \frac{1}{Q-1} \sum_{i=1}^Q (f_{m_i} - \bar{\mu}_k)^2, \quad (4.10)$$

где Q — общее число обучающих образов класса k , f_{m_i} — яркость пикселя этого класса из m -й обучающей выборки.

Предполагая, что функция вероятности любого класса может быть аппроксимирована нормальной функцией плотности, вместо всей гистограммы в памяти ЭВМ можно хранить только математические ожидания и дисперсии каждого класса $\bar{\mu}_k$ и $\bar{\sigma}_k^2$. При двух спектральных каналах двумерная функция распределения для каждого класса может быть оценена путем записи частот повторения всех возможных пар значений яркостей одного и того же пикселя в двух каналах: каждая пара состоит из значения f_{ij1} первого канала и значения f_{ij2} — второго. Результат будет представлять собой двумерную гистограмму. Но, как и в одномерном случае, для экономии памяти ЭВМ можно предположить возможность параметрического представления функции распределения вероятностей нормальной функцией плотности вероятности. Двумерная нормальная функция плотности вероятностей имеет вид

$$w_2(f_{ij1}, f_{ij2}) = A \exp B \left[\frac{(f_{ij1} - \mu_{1k})^2}{\sigma_{1k}^2} + \frac{(f_{ij2} - \mu_{2k})^2}{\sigma_{2k}^2} - 2\rho_{12k} \frac{(f_{ij1} - \mu_{1k})(f_{ij2} - \mu_{2k})}{\sigma_{1k}^2 \sigma_{2k}^2} \right], \quad (4.11)$$

где μ_{1k} и $\sigma_{1k}^2 = R_{11k}$, μ_{2k} и $\sigma_{2k}^2 = R_{22k}$ — математические ожидания и дисперсии яркости пиксела 1-го и 2-го каналов (для класса k); $A = (2\pi\sigma_{1k}\sigma_{2k}\sqrt{1-\rho_{12k}^2})^{-1}$, $B = -[2(1-\rho_{12k}^2)]^{-1}$, $\rho_{12k} = \rho_{21k} = R_{12k}/\sigma_{1k}\sigma_{2k}$

$$R_{12k} = R_{21k} = \langle (f_{ij1} - \mu_{1k})(f_{ij2} - \mu_{2k}) \rangle \quad (4.12)$$

— коэффициент корреляции между яркостями пиксела в каждом канале (для класса k). На практике R_{12k} неизвестен и должен оцениваться по обучающим выборкам:

$$\bar{R}_{12k} = \frac{1}{Q-1} \sum_{m=1}^Q [f_{m1} - \bar{\mu}_{1k}][f_{m2} - \bar{\mu}_{2k}], \quad (4.13)$$

где f_{m1} и f_{m2} — значения яркости пикселей класса k из m -й обучающей выборки.

Если параметры $\bar{\mu}_{1k}$, $\bar{\mu}_{2k}$, \bar{R}_{11k} , \bar{R}_{22k} , \bar{R}_{12k} записаны в памяти ЭВМ для каждого класса (всего пять параметров для каждого класса), функции распределения вероятностей для данных могут быть вычислены по формуле (4.11). Для ее упрощения можно использовать векторную (матричную) форму записи. Для общего случая n каналов обозначим вектор данных (i, j) -го пиксела через \mathbf{f}_{ij} , вектор математического ожидания для k -го канала — через $\boldsymbol{\mu}_k$, корреляционную матрицу — через \mathbf{R}_k :

$$\mathbf{f}_{ij} = \begin{bmatrix} f_{ij1} \\ f_{ij2} \\ \vdots \\ f_{ijn} \end{bmatrix}, \quad \boldsymbol{\mu}_k = \begin{bmatrix} \mu_{1k} \\ \mu_{2k} \\ \vdots \\ \mu_{nk} \end{bmatrix}, \quad \mathbf{R}_k = \begin{bmatrix} R_{11k} & R_{12k} & R_{13k} & \dots & R_{1nk} \\ R_{21k} & R_{22k} & R_{23k} & \dots & R_{2nk} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{n1k} & R_{n2k} & R_{n3k} & \dots & R_{nnk} \end{bmatrix}, \quad (4.14)$$

n -мерную нормальную функцию плотности вероятностей — через

$$w_n(\mathbf{f}_{ij}) = (2\pi)^{-n/2} |\mathbf{R}_k|^{-1/2} \exp[-S(\mathbf{f}_{ij} - \boldsymbol{\mu}_k)^T \mathbf{R}_k^{-1} (\mathbf{f}_{ij} - \boldsymbol{\mu}_k)], \quad (4.15)$$

где $|R_k|$ — определитель ковариационной матрицы, R_k^{-1} — обратная матрица, $(f_{ij} - \mu_k)^T$ — транспонированный вектор $(f_{ij} - \mu_k)$.

Таким образом, при справедливости параметрической гипотезы нормальном законе распределения яркости пикселей можно описывать классы образов с помощью векторов математических ожиданий и корреляционных матриц. Опыт показывает, что методы классификации, основанные на нормальном законе распределения, удовлетворительно работают даже при значительных отклонениях от нормальности.

Однако при использовании предположения о нормальности следует уделить внимание двум вопросам. Во-первых, нужно подбирать адекватные обучающие выборки, позволяющие оценивать математические ожидания и корреляционные матрицы каждого класса. Если при классификации используются данные n спектральных каналов, теоретически минимальное число обучающих образов (это могут быть одиночные пиксели) для каждого класса равно $n + 1$. При меньшем числе обучающих образов корреляционная матрица будет вырожденной (ее определитель равен 0, и обратная матрица не существует), что делает невозможным вычисление (4.15). На практике необходимо иметь 10 и более обучающих образов, чтобы получить надежные оценки параметров классов. Второй вопрос касается случаев, когда закон распределения явно многомодальный. Распространенное практическое решение — представление многомодальной функции распределения в виде суммы нормальных законов.

4.2. Метод максимального правдоподобия

Пусть спутниковое сканерное изображение земной поверхности, полученное в n спектральных каналах, представляет собой совокупность пикселей, причем яркость (i, j) -го пикселя описывается вектором из n компонент — значениями яркости в каждом канале. Требуется отнести этот пиксел к одному из K классов. Для этого предварительно создается обучающая выборка — спутниковое изображение, на котором по наземным данным определены участки, отвечающие различным классам (например, водоемы, сельхозпосевы, городские и сельские застройки и т.п.). Пользователь указывает границы этих классов, происходит процесс обучения, т.е. компьютер оценивает вектор математического ожидания и корреляционную матрицу для каждого из классов.

Далее производится обработка интересующего нас спутникового изображения. Для каждого пикселя компьютером определяется вектор измерений f_{ij} и в каждом классе k формируются условные плотности вероятностей $p(f_{ij}/C_k)$. Для этого в многомерный нормальный закон распределения (4.15) вместо X_m подставляется f_{ij} , а в качестве μ_k и R_k — за-

даваемые условия, т.е. упомянутые оценки вектора математического ожидания и корреляционной матрицы. Таким образом, для каждого пиксела формируется K условных плотностей вероятностей, которые сравниваются с некоторым порогом Q_{kl} согласно (4.7)–(4.8). Сравнение позволяет определить, какая из гипотез наиболее правдоподобна. В этом и состоит метод максимального правдоподобия. Используя (4.15), получаем, что отношение правдоподобия L_{kl} для классов k и l

$$L_{kl} = \frac{f_{ij}/C_k}{p(f_{ij}/C_k)} = \frac{|\bar{\mathbf{R}}_k|^{-1/2} \exp[-(1/2)(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)^t \bar{\mathbf{R}}_k^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)]}{|\bar{\mathbf{R}}_l|^{-1/2} \exp[-(1/2)(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_l)^t \bar{\mathbf{R}}_l^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_l)]}.$$

Вместо отношения правдоподобия L_{kl} можно использовать любую монотонную функцию от него, например натуральный логарифм:

$$\ln L_{kl} = \frac{1}{2} \ln \frac{|\bar{\mathbf{R}}_l|}{|\bar{\mathbf{R}}_k|} + \frac{1}{2} [(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)^t \bar{\mathbf{R}}_k^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)] - \frac{1}{2} [(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_l)^t \bar{\mathbf{R}}_l^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_l)].$$

В правой части в (4.7), (4.8) в этом случае будет стоять $q_{kl} = \ln Q_{kl}$. Первый член $(1/2) \ln[|\bar{\mathbf{R}}_l|/|\bar{\mathbf{R}}_k|]$ и множитель $1/2$ в выражении для $\ln L_{kl}$ могут быть включены в q_{kl} . Таким образом, задача выяснения, к какому классу относится (i, j) -й пиксел, сводится к сравнению квадратичных форм

$$[(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)^t \bar{\mathbf{R}}_k^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)] \text{ и } [(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_l)^t \bar{\mathbf{R}}_l^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_l)].$$

Для простоты нередко полагают, что $q_{kl} = 0$. В этом случае считается, что (i, j) -й пиксел относится к k -му классу, если для него эта квадратичная форма меньше, т.е. если компоненты вектора \mathbf{f}_{ij} более тесно группируются вокруг $\bar{\boldsymbol{\mu}}_k$.

Квадратичная форма $[(\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)^t \bar{\mathbf{R}}_k^{-1} (\mathbf{f}_{ij} - \bar{\boldsymbol{\mu}}_k)]$ характеризует расстояние между объектом \mathbf{f}_{ij} и средним значением $\bar{\boldsymbol{\mu}}_k$ и часто применяется в задачах распознавания изображений. Для вектора \mathbf{f}_{ij} , состоящего из двух компонент, т.е. для двух спектральных каналов, квадратичная форма имеет вид

$$M_k = \frac{1}{(1 - \bar{\rho}_{12k}^2)} \left[\frac{(f_{ij1} - \bar{\mu}_{1k})^2}{\bar{\sigma}_{1k}^2} + \frac{(f_{ij2} - \bar{\mu}_{2k})^2}{\bar{\sigma}_{2k}^2} - 2\bar{\rho}_{12k} \frac{(f_{ij1} - \bar{\mu}_{1k})(f_{ij2} - \bar{\mu}_{2k})}{\bar{\sigma}_{1k} \bar{\sigma}_{2k}} \right]. \quad (4.16)$$

Распознавание методом максимального правдоподобия (МП) требует достаточно большой объем статистических данных и позволяет успешно решать сложные задачи. При справедливости предположения о нормальном законе распределения вектора f_{ij} метод МП обеспечивает оптимальное распознавание. Однако, как можно заключить из сказанного, этот метод требует достаточно большого числа операций и является сравнительно медленным. При его использовании для обработки изображений желательно учитывать в спектральных каналах не только яркости каждого одиночного пиксела (попиксельная классификация), но и его окружение. К сожалению, такой учет приводит к еще большему замедлению процесса распознавания и на практике почти не применяется. Поэтому в некоторых случаях рассматриваемый вариант метода МП проигрывает в качестве распознавания более простым методам, которые принимают во внимание взаимные связи соседних пикселей. Причиной неверной классификации могут быть также завышенные значения R_{mnk} .

Результат попиксельной классификации выгоревших участков леса (исходное изображение на рис. 4.2) методом МП можно видеть на рис. 4.6. При обработке изображения использовалась обучающая выборка из 12 классов, как в методе ISODATA. Результаты классификации несколько хуже, чем в методах ISODATA и сегментации путем наращивания областей.

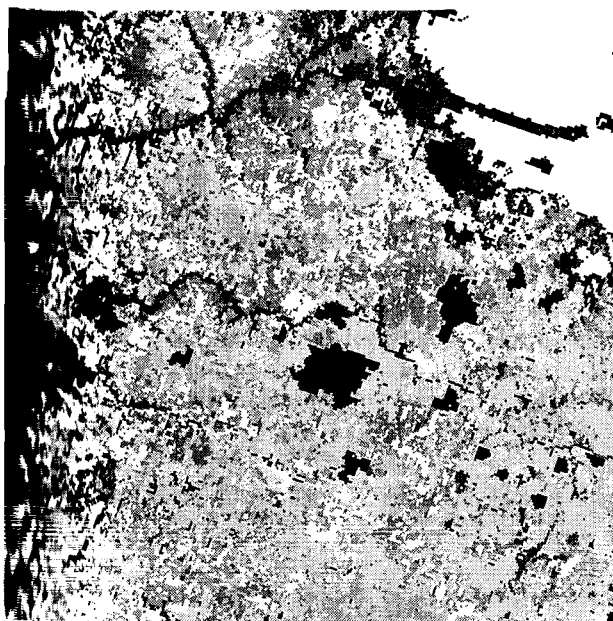


Рис. 4.6. Результат классификации по методу максимального правдоподобия

4.6. ПРОСТЫЕ МЕТОДЫ, ОСНОВАННЫЕ НА ГРУППИРОВАНИИ

Квадратичную форму $[(f_{ij} - \bar{\mu}_k)^t \bar{R}_k^{-1} (f_{ij} - \bar{\mu}_k)]$ можно упростить, если положить $R_{mpk} = 0$, $m \neq p$, т.е. не учитывать корреляцию яркостей пиксела в спектральных каналах. При этом в (4.16) все коэффициенты $r_{12k} = 0$, $k = 1, 2, \dots, n$. В квадратичной форме (4.16) останутся только два члена, множитель перед квадратными скобками обратится в 1. В общем случае получаем выражение

$$r_k^2 = \sum_{m=1}^n (f_{ijm} - \bar{\mu}_{mk})^2 / \bar{\sigma}_{mk}^2, \quad (4.17)$$

где m — номер спектрального канала, n — общее число каналов. При этом считается, что пиксел относится к тому классу, для которого меньше значение r_k^2 . Однако точность определения r_k^2 существенно зависит от точности оценивания $\bar{\sigma}_{mk}^2$, как и при методе МП. Дальнейшее упрощение приводит к *методу минимальных расстояний (минимальных дистанций)*, когда решение о том, к какому классу относится пиксел, принимается на основании сравнения расстояний между пикселами и средними значениями в пространстве яркостей:

$$r_k^2 = \sum_{m=1}^n (f_{ijm} - \bar{\mu}_{mk})^2. \quad (4.18)$$

Этот самый простой и быстрый метод классификации требует минимальных сведений о классах, но уступает методу МП в точности. Он не учитывает окружения пиксела, поэтому, как правило, уступает и методам кластерного анализа, и методам сегментации по наращиванию областей и выделению границ. Метод минимальных расстояний основан на евклидовой метрике и может рассматриваться безотносительно к нормальному закону распределения.

Результат классификации методом минимальных расстояний можно видеть на рис. 4.7. Использовалась та же обучающая выборка, что и в методах МП и ISODATA. Метод минимальных расстояний легко выделяет свежие, но пропускает старые пожарища, не относя к этим классам другие объекты.

На практике применяют еще один простой метод классификации, связанный с нормальным законом, — *метод параллелепипедов*, называемый иногда *методом прямоугольников*. Он использует сведения о классах в виде векторов средних значений яркости $\bar{\mu}_k$ и векторов дисперсии

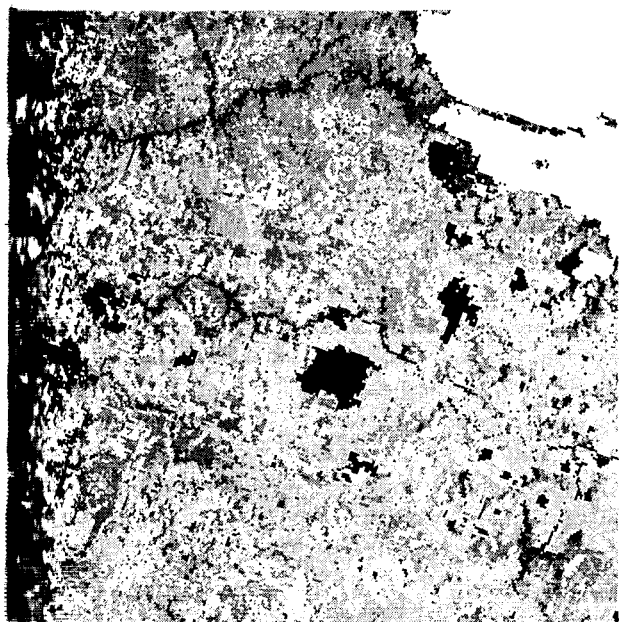


Рис. 4.7. Результат классификации по методу минимальных расстояний

$\bar{\mu}_{mk}$, полученные в процессе обучения. Известно, что для нормально распределенной случайной величины 95,4% ее значений лежат в пределах отклонений от среднего значения, меньших 2σ . Рассмотрим случай двух спектральных компонентов. На рис. 4.8 по оси абсцисс отложены значения яркости f_1 в 1-м канале, по оси ординат — значения яркости f_2 во

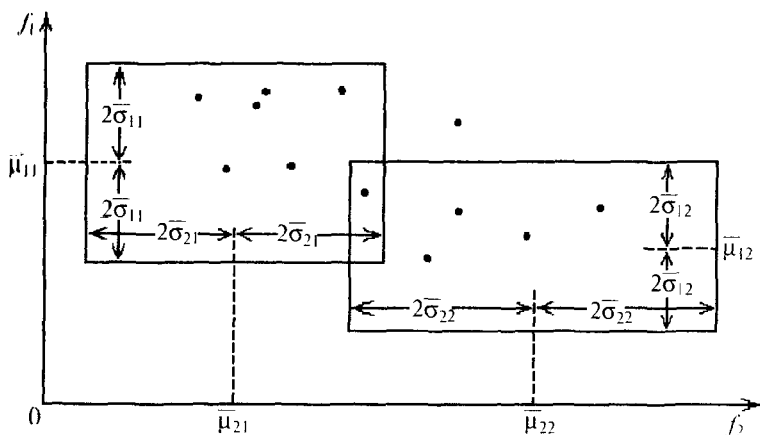


Рис. 4.8. Результат классификации по методу параллелепипедов

2-м канале. Здесь же указаны средние значения яркости μ_{f_1} и μ_{f_2} для первого и второго классов в 1-м канале, μ_{f_1} и μ_{f_2} для первого и второго классов во 2-м канале. От значения средней яркости в обе стороны отложены значения среднеквадратического отклонения σ_{f_1} , σ_{f_2} , σ_{f_1} , σ_{f_2} для первого и второго классов в каждом из каналов. В двумерном случае получились прямоугольники, в трехмерном и более — параллелепипеды в соответствующем пространстве. Если компоненты вектора яркости f_{ij1} и f_{ij2} попадают в левый прямоугольник, принимается решение о принадлежности пиксела к первому классу, если они попадают в правый прямоугольник — то ко второму. Прямоугольники могут частично перекрываться, что особенно характерно для областей, примыкающих к углам. В этом случае возникает неопределенность.

Метод параллелепипедов относится к числу быстрых методов классификации, однако по качеству распознавания уступает большинству из рассмотренных. Его целесообразнее всего использовать для предварительной классификации. Он наиболее эффективен при нормальном законе, так как для него выполняется упомянутое «правило двух сигма».

4.7. НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ КЛАССИФИКАЦИИ

Задача непараметрической классификации возникает, когда вид функций распределения $w_n(\mathbf{f}_{ij})$ для совокупности классов неизвестен. Она состоит в том, чтобы по классифицированным обучающим выборкам $\{\mathbf{f}_{ij}\}$ оценить плотности распределений $w_n(\mathbf{f}_{ij})$ и/или сделать необходимые выводы о принадлежности вектора \mathbf{f}_{ij} тому или иному классу. Априорная информация, используемая при синтезе непараметрических систем классификации, скорее носит качественный характер, чем количественный. В непараметрических методах главный упор делается не на оптимизацию процесса классификации, а на обеспечение нечувствительности к условиям функционирования. Синтез непараметрических классификаторов проводится преимущественно эвристическими методами.

В настоящее время при непараметрической классификации изображений в основном используют метод многомерных гистограмм, метод Парзена, метод разложений по базисным функциям, метод полигонов Смирнова, различные методы локального оценивания, а также некоторые методы нелинейного оценивания. Остановимся на методе многомерных гистограмм.

При пороговой сегментации (см. рис. 4.1) было использовано свойство одномерной гистограммы (см. рис. 3.7) иметь моду (максимум)

в некоторой области пространства измерений, в которой и стремится расположиться большая часть значений яркости. Иными словами, значения яркости объектов имеют тенденцию группироваться вблизи моды, что справедливо и для многомерных гистограмм (рис. 4.9). Это позволяет провести классификацию многоспектральных изображе-

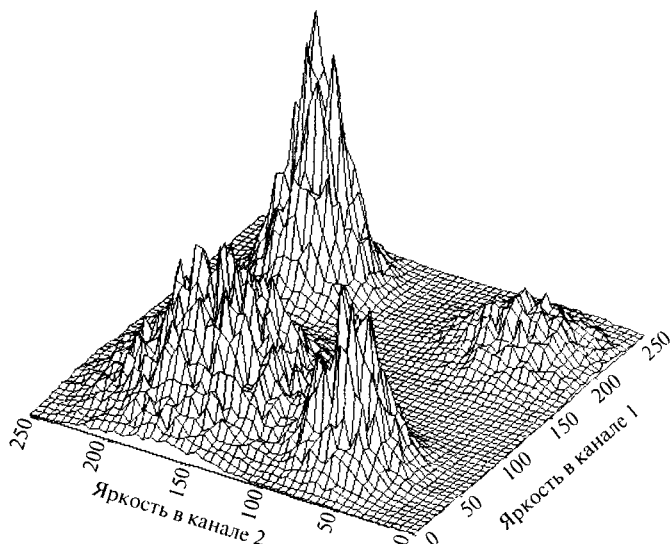


Рис. 4.9. Двумерная гистограмма

ний. Метод классификации по максимумам многомерных гистограмм относится к числу непараметрических и как таковой не требует обучения, однако для интерпретации результатов классификации необходимы априорные данные или экспертные оценки.

К числу непараметрических методов локального оценивания можно отнести методы минимальных расстояний и параллелепипедов, как и классификацию на основании вычисления квадратичной формы вида (4.16), потому что они могут использоваться при произвольных и неизвестных заранее законах распределения вектора яркости пиксела.

В заключение кратко обобщим результаты классификации изображения на рис. 4.2 в трех спектральных каналах: красном участке видимого диапазона, в ближнем ИК- и в тепловом канале. Все применявшиеся методы — ISODATA, методы наращивания областей, выделения контуров, максимального правдоподобия, параллелепипедов, минимальных расстояний, максимумов многомерных гистограмм — позволили в той или иной степени выделить разновозрастные пожарища площадью свыше 2000 га или бо-

лес чем 4×4 пиксела (пространственное разрешение составляло 1,1 км). Лучшие результаты для выделения пожарищ дал метод наращивания областей, учитывающий пространственную структуру объекта. Однако эта интерактивная процедура требует значительных затрат времени и участия эксперта. Аналогичные результаты получены при использовании алгоритма ISODATA с обучающей выборкой при условии, что заданы большое количество итераций и высокий порог сходимости δ .

Метод МП достаточно полно выделяет пиксеты, относящиеся к разновозрастным пожарищам, но вместе с ними относит к этим классам края облаков, берега рек и т.п. В использованном варианте метода не учитывалась связность объектов и производилась индивидуальная классификация каждого пиксела.

Метод классификации по максимумам трехмерных гистограмм позволил выделить все крупные пожарища, но разновозрастные пожарища были объединены в один класс. Этот метод не относится к числу быстрых, так как на построение многомерной гистограммы требуется значительное время, сравнимое с временем обработки методом МП.

Метод минимальных расстояний легко выделяет свежие, но пропускает некоторые старые пожарища, не относя к этим классам другие объекты. По скорости обработки это самый быстрый алгоритм. Быстрым, но самым неточным оказался метод параллелепипедов. Метод, основанный на выделении контуров, не учитывал пожарища с малой протяженностью контура.

4.8. АНАЛИЗ ГЛАВНЫХ КОМПОНЕНТОВ

Совокупность признаков, используемых при распознавании образов, должна в наибольшей степени отражать те свойства объектов, которые важны для распознавания. От размерности p признакового пространства в значительной степени зависят вычислительная сложность процедур обучения и принятия решения, достоверность распознавания и затраты времени.

При отображении на мониторе ЭВМ результатов многоспектральных дистанционных исследований возникает еще одна проблема: нам доступны три цвета (RGB), которые можно присвоить как условные трем спектральным каналам, в то время как число каналов может превышать эту цифру (7 каналов сканера ТМ спутника LANDSAT, 5 каналов сканера AVHRR NOAA). И в этом случае важно так преобразовать спутниковую информацию, чтобы ее можно было адекватно представить на экране монитора.

Кроме того, сокращение числа признаков иногда способно увеличить расстояние между признаками и упростить процедуру распознавания.

Первоначальный набор признаков формируется до начала распознавания из числа доступных измерению характеристик объекта g_1, g_2, \dots, g_p , отражающих существенные для распознавания свойства. На следующем этапе из первоначального можно сформировать новый набор h_1, h_2, \dots, h_q , состоящий из меньшего числа признаков: $q < p$.

Уменьшение числа признаков снижает затраты, но может привести к снижению достоверности распознавания. Таким образом, требования минимума общей размерности задачи и максимума достоверности распознавания оказываются противоречивыми, поэтому для уменьшения числа признаков необходимо разумно выбрать размерность признакового пространства или выполнение правильного, без заметного уменьшения информативности, преобразования совокупности данных.

В качестве критерия правильности выбора новых признаков используют среднеквадратическую ошибку аппроксимации признаков $\{g_1, g_2, \dots, g_p\}$ с помощью $\{h_1, h_2, \dots, h_q\}$, среднее межклассовое расстояние, внутриклассовый разброс наблюдений, энтропию одного класса относительно другого и т.д.

Наиболее широкое распространение получили методы линейного преобразования \mathbf{A} исходного пространства признаков $\mathbf{G} = (g_1, g_2, \dots, g_p)$ в новое пространство $\mathbf{H} = (h_1, h_2, \dots, h_q)$:

$$\mathbf{H} = \mathbf{A}\mathbf{G}. \quad (4.19)$$

Преобразование (4.19) позволяет перейти от одной координатной системы, в которой признаки g_i , например компоненты векторов яркости пикселей \mathbf{f}_{ij} , коррелированы, к другой, где они не коррелированы. В этом случае процедура носит название *преобразования Карунена—Лоэва*. В качестве столбцов матрицы преобразования \mathbf{A} выбирают собственные векторы общей корреляционной матрицы \mathbf{R} распознаваемых совокупностей. Сама корреляционная матрица \mathbf{R}^* становится диагональной с собственными числами λ_i на диагонали:

$$\mathbf{R}^* = \mathbf{A}^t \mathbf{R} \mathbf{A} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

После этого преобразования отбирают q ($q < p$) новых признаков, соответствующих тем собственным числам λ_i матрицы \mathbf{R}^* , которые оказывают наибольшее влияние на значения выбранных критериев близости $\{g_1, g_2, \dots, g_p\}$ и $\{h_1, h_2, \dots, h_q\}$. В частности, при использовании критерия наилучшей аппроксимации межклассового расстояния нецелесообразно оставить те признаки, которые соответствуют максимальным собственным числам λ_j , т.е. выбрать *главные компоненты* после преобразования \mathbf{A} .

Кроме этого метода выбора признаков, существуют и другие, например выбор по критерию совместной минимизации внутриклассового разброса наблюдений и максимизации межклассового расстояния. Пусть наблюдение Земли со спутника производится в двух спектральных каналах, т.е. вектор яркости \mathbf{f}_j содержит два компонента. Корреляционная матрица яркостей пикселей имеет вид

$$\mathbf{R} = \begin{vmatrix} 2,6 & 1,2 \\ 1,2 & 1,3 \end{vmatrix}, \quad (4.20)$$

где $\sigma_1^2 = 2,6$ и $\sigma_2^2 = 1,3$ — дисперсия яркости в 1-м и во 2-м канале, $R_{12} = R_{21} = 1,2$ — коэффициент корреляции между значениями яркости в каналах.

Собственные числа матрицы \mathbf{R} можно найти из решения уравнения

$$\begin{vmatrix} 2,6 - \lambda & 1,2 \\ 1,2 & 1,3 - \lambda \end{vmatrix} = 0,$$

т.е. $\lambda^2 - 3,9\lambda + 1,94 = 0$, откуда $\lambda_1 = 3,31$, $\lambda_2 = 0,58$, а преобразованная матрица

$$\mathbf{R}^* = \begin{vmatrix} 3,31 & 0 \\ 0 & 0,58 \end{vmatrix}.$$

Собственные векторы \mathbf{g}_i матрицы \mathbf{R} являются решениями уравнений $[\mathbf{R} - \lambda_i \mathbf{I}]\mathbf{g}_i = 0$, $i = 1, 2$, где \mathbf{I} — единичная матрица. Эти уравнения имеют нетривиальные решения, так как их определители равны нулю. Потребуем также, чтобы компоненты собственных векторов $g_{1i}^2 + g_{2i}^2 = 1$. Векторы \mathbf{g}_1 и \mathbf{g}_2 с компонентами $g_{11} = 0,86$, $g_{21} = -0,51$ и $g_{12} = 0,51$, $g_{22} = 0,86$ образуют столбцы матрицы \mathbf{A} .

Выражение (4.19) принимает вид

$$\begin{vmatrix} s_1 \\ s_2 \end{vmatrix} = \begin{vmatrix} 0,86 & 0,51 \\ -0,51 & 0,86 \end{vmatrix} \begin{vmatrix} f_1 \\ f_2 \end{vmatrix},$$

где s_1 , s_2 — новые значения яркости пикселей в каналах, f_1 , f_2 — старые значения. Дисперсия s_1 равна $\lambda_1 = 3,31$, дисперсия s_2 равна $\lambda_2 = 0,58$; на s_1 приходится 84,9% суммарной дисперсии — это главный компонент.

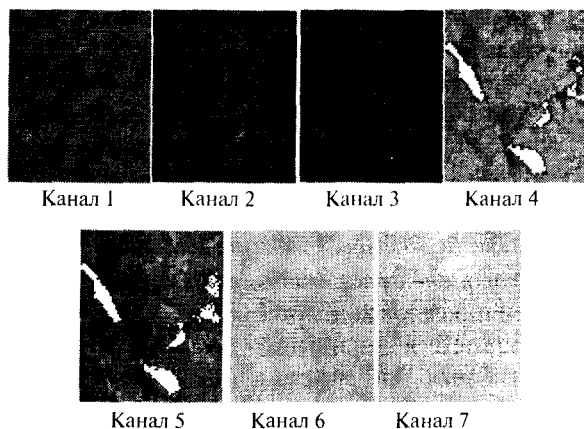


Рис. 4.10. Изображения поверхности в семи спектральных каналах

На рис. 4.10 показаны изображения участка поверхности, полученные с помощью спутника LANDSAT, сканер TM в семи спектральных каналах, т.е. вектор яркости f_{ij} содержит семь компонентов: канал 1 — интервалу длин волн 0,45–0,52 мкм; канал 2 — интервалу 0,52–0,60 мкм; канал 3 — интервалу 0,63–0,69 мкм; канал 4 — интервалу 0,76–0,90 мкм; канал 5 — интервалу 1,55–1,75 мкм; канал 6 — интервалу 10,4–12,5 мкм; канал 7 — интервалу 2,08–2,35 мкм. В этом примере использованы материалы из пакета прикладных программ IDRISI.

Корреляционная матрица яркости пикселей компонентов вектора f_{ij} имеет вид

$$R = \begin{vmatrix} 19,36 & 11,72 & 17,74 & 22,99 & 42,24 & 8,58 & 19,01 \\ 11,72 & 9,47 & 12,55 & 33,96 & 40,96 & 5,65 & 15,64 \\ 17,74 & 12,55 & 20,20 & 23,83 & 48,88 & 9,05 & 21,74 \\ 22,99 & 33,96 & 23,83 & 606,05 & 370,12 & 5,60 & 82,75 \\ 42,24 & 40,96 & 48,88 & 370,12 & 329,80 & 21,38 & 96,77 \\ 8,58 & 5,65 & 9,05 & 5,06 & 21,38 & 8,50 & 10,47 \\ 19,01 & 15,64 & 21,74 & 82,75 & 96,77 & 10,47 & 35,30 \end{vmatrix}. \quad (4.21)$$

На главной диагонали матрицы (4.21) записаны значения дисперсии компонентов вектора яркости изображения. Так, в первом столбце 19,36 — это дисперсия яркости в канале 1. В четвертом столбце

606,05 — дисперсия яркости в канале 4, в пятом столбце 329,80 — дисперсия яркости в канале 5. Напомним, что дисперсия характеризует контраст изображения. Как видно из рис. 4.10, изображения в каналах 4 и 5 наиболее контрастны. Числа вне главной диагонали — это коэффициенты корреляции между компонентами вектора яркости изображения. Число $R_{12} = R_{21} = 11,72$ в первом и втором столбцах — ненормированный коэффициент корреляции между яркостями пикселей каналов 1 и 2, число $R_{27} = R_{72} = 15,64$ в последней строке второго столбца и в седьмом столбце — ненормированный коэффициент корреляции между яркостями пикселей каналов 2 и 7.

Решение уравнения, подобного (4.19), в данном случае даст семь собственных значений матрицы (4.21): $\lambda_1 = 890,14$; $\lambda_2 = 114,83$; $\lambda_3 = 15,53$; $\lambda_4 = 3,85$; $\lambda_5 = 1,79$; $\lambda_6 = 1,78$; $\lambda_7 = 0,76$. В качестве примера приведем также первый и четвертый столбцы матрицы преобразования A , которые являются собственными векторами матрицы (4.21):

$$C_1 = (0,05412; 0,06179; 0,06001; 0,80203; 0,56966; 0,02218; 0,14624),$$

$$C_4 = (-0,20055; -0,13330; -0,20633; 0,02439; -0,01599; 0,94794; 0,00000).$$

Результат преобразования многоспектрального изображения на рис. 4.10 с помощью процедуры Карунена—Лоэва можно видеть на рис. 4.11. В данном случае уже нельзя говорить о том, к какому спек-

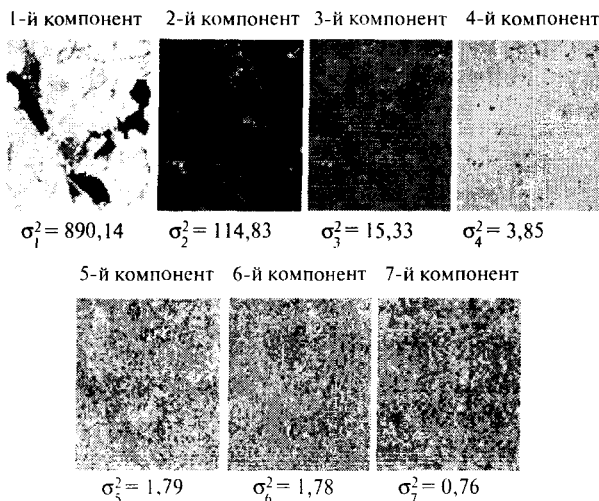


Рис. 4.11. Результат преобразования Карунена—Лоэва

рильному каналу относятся компоненты преобразованного изображения. Наибольшей дисперсией обладают 1-й и 2-й компоненты на рис. 4.11; на 1-й приходится 86,53% суммарной дисперсии, на 2-й — 11,16%, на 3-й — 1,51%, на 4-й — 0,37%, на 5-й — 0,17%, на 6-й — 0,17%, на 7-й — 0,07%. Первые два компонента являются *главными компонентами* многоспектрального изображения на рис. 4.10. Все компоненты на рис. 4.11 не коррелированы между собой. Интересно отметить, что преобразование Карунена—Лоэва в данном случае даст отрицательные значения яркости пикселей 5–7-го компонента, на рис. 4.11 эти компоненты показаны после дополнительного преобразования, сделавшего яркость положительной. Сравнение их с рис. 2.12 показывает, что эти компоненты представляют собой случайные поля с практически некоррелированными значениями яркости, т.е. шум.

4.9. ПРИМЕНЕНИЕ НЕЙРОКОМПЬЮТЕРОВ

Нейрокомпьютеры — мощный инструмент в задачах распознавания изображений — позволяют стандартным способом решать многие нестандартные задачи. Они особенно эффективны там, где требуется поддobie человеческой интуиции, — для распознавания образов (распознавание сложных изображений, чтение текстов и т.п.). Именно для таких задач обычно трудно формализовать набор признаков и создать явный алгоритм. Основной структурной единицей нейрокомпьютера является *формальный нейрон*, суммирующий входные сигналы с определенными весовыми коэффициентами и выдающий результирующий сигнал на входы других нейронов и на выход нейронной сети. Обучение нейрокомпьютера сводится к подбору весовых коэффициентов для получения требуемого результата.

В основе используемых представлений лежит идея о том, что нейроны можно моделировать довольно простыми автоматами, а вся сложность нейрокомпьютера, гибкость его функционирования и другие важнейшие качества определяются *связями между нейронами*. Каждая связь представляется как совсем простой элемент, служащий для передачи сигнала.

Особенностью нейрокомпьютерного подхода является использование обучения вместо программирования. Труд программиста заменяется трудом учителя (тренера). Программист предписывает ЭВМ все детали работы, учитель — создает «образовательную среду», к которой приспособливается нейрокомпьютер. Обучение обычно строится на основе задачника — набора примеров с заданными ответами. Эти примеры предъявляются системе. Нейроны получают по входным связям

сигналы — условия примера, преобразуют их, несколько раз обмениваются преобразованными сигналами и наконец выдают ответ — также набор сигналов. Отклонение от правильного ответа штрафуются. Обучение состоит в минимизации штрафа как функции связей.

Для описания алгоритмов и устройств в нейроинформатике разработана специальная «схемотехника», в которой элементарные устройства — сумматоры, синапсы, нейроны и т.п. — объединяются в сети, предназначенные для решения задач. Схемотехника представляет собой особый язык для описания нейронных сетей. При программной и аппаратной реализации описания переводятся на языки другого уровня.

Самый важный элемент нейросистем — *адаптивный сумматор*, который вычисляет скалярное произведение вектора входного сигнала x и вектора настраиваемых параметров \tilde{a} .

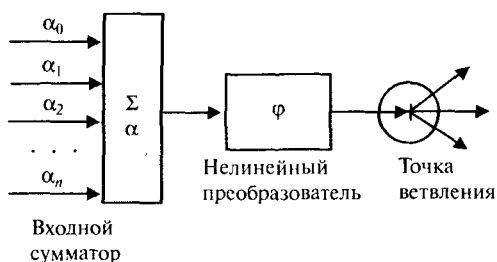


Рис. 4.12. Формальный нейрон

Стандартный формальный нейрон состоит из входного сумматора, нелинейного преобразователя и точки ветвления на выходе (рис. 4.12). *Нелинейный преобразователь сигнала* получает скалярный входной сигнал x и переводит его в $\varphi(x)$. *Точки ветвления* служит для рассылки сигнала по несколь-

ким адресам: она принимает скалярный входной сигнал x и передает его всем своим выходам.

Для создания нейронных сетей применяют несколько стандартных архитектур, из которых строят большинство используемых сетей. Первый вопрос, с которым приходится при этом сталкиваться, — согласованная работа различных нейронов во времени. Для программных имитаторов нейронных сетей на цифровых ЭВМ этот вопрос снимается автоматически. Для других способов реализации он весьма важен.

Можно выделить две базовые архитектуры — *слоистые и полносвязные сети*. В слоистых сетях нейроны расположены в несколько слоев (рис. 4.13). Нейроны слоя 1 получают входные сигналы, преобразуют их и через точки ветвления передают нейронам слоя 2. Далее срабатывает слой 2, и так далее до слоя k , который выдает выходные сигналы для интерпретатора и пользователя. Число нейронов в каждом слое может быть любым и никак заранее не связано с числом нейронов в других слоях. Особо распространены трехслойные сети, в которых каждый слой имеет свое наименование: входной, скрытый, выходной.

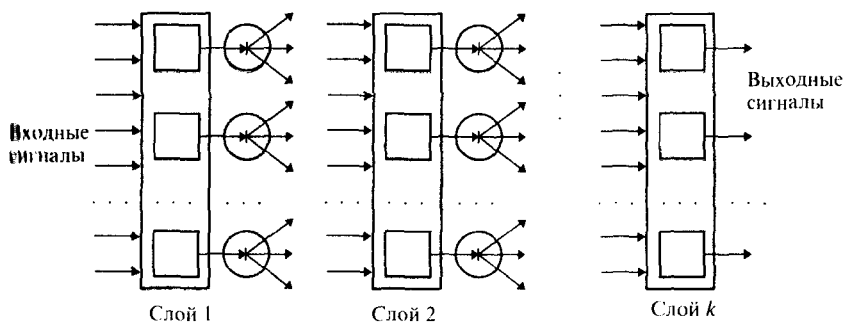


Рис. 4.13. Схема слоистой нейронной сети

В полносвязных сетях каждый нейрон передает свой выходной сигнал остальным нейронам, включая самого себя. Выходными сигналами сети могут быть все или некоторые выходные сигналы нейронов после нескольких тактов функционирования сети. Все входные сигналы подаются всем нейронам. Для полносвязной сети входной сумматор нейрона фактически распадается на два: первый вычисляет линейную функцию от входных сигналов сети, второй — линейную функцию от выходных сигналов других нейронов, полученных на предыдущем шаге.

Функция активации нейронов (характеристическая функция) ϕ — нелинейное преобразование выходного сигнала сумматора (см. рис. 4.12) — может быть одной и той же для всех нейронов, в этом случае сеть называют *однородной (гомогенной)*. Если же ϕ зависит еще от одного или нескольких параметров, значения которых меняются от нейрона к нейрону, то сеть называют *неоднородной (гетерогенной)*.

Составление сети из нейронов стандартного вида не обязательно. Слоистая или полносвязная архитектура не налагает существенных ограничений на участвующие в ней элементы. Единственное жесткое требование, предъявляемое архитектурой к элементам сети, — это соответствие размерности вектора входных сигналов элемента (она определяется архитектурой) числу его входов.

Если полносвязная сеть функционирует заданное число тактов k до получения ответа, то ее можно представить как частный случай k -слоистой сети, все слои которой одинаковы, причем каждый из них соответствует такту функционирования полносвязной сети. Существенное различие между полносвязной и слоистой сетями возникает тогда, когда число тактов функционирования заранее не ограничено — слоистая сеть так работать не может.

Важным моментом в организации нейронной сети является *обучение*. Если требования к нейронной сети сформулировать как задачу

минимизации некоторой функции — оценки, зависящей от части сигналов (входных, выходных) и от параметров сети, то обучение можно рассматривать как оптимизацию и строить соответствующие алгоритмы, программное обеспечение и, наконец, устройства.

Выбор методов минимизации оценки — сложная и в итоге практически наиболее важная проблема, от успеха реализации которой зависит скорость обучения. В настоящее время выработаны достаточно эффективные методы минимизации оценки, позволяющие за приемлемое время осуществлять обучение нейροкомпьютера распознаванию образов.

В качестве примера рассмотрим задачу навигации искусственного спутника Земли на геостационарной орбите. Использовалась программная имитация на ЭВМ двухслойной нейронной сети из десяти 10-входных нейронов. Изображение земной поверхности, полученное на матричном фотоприемнике форматом 600×600 пикселей с углом зрения $1^\circ 30'$, обрабатывается нейрокомпьютером. Требовалось четко разделить изображения различных объектов (острова, облака, океан и т.д.) и выделить опорные точки, по которым производится оценка углового положения спутника относительно его внутренней системы координат (ось X лежит в плоскости орбиты спутника, ось Z направлена по местной вертикали в сторону центра Земли). Наблюдение осложнялось угловыми колебаниями спутника и появлением облаков (ставилась задача обеспечить уверенное распознавание опорных точек на береговой линии на 70–80% закрытой случайным образом расположенной облачностью).

Обучение нейронной сети сводилось к подбору весовых коэффициентов на входах нейронов для получения на выходе сети нужных результатов при различных отклонениях оси Z от направления на центр Земли и различных покрытиях объектов облачностью. Точность работы нейрокомпьютера была исследована методами имитационного моделирования на ЭВМ. Координаты опорных точек, определенные нейрокомпьютером, сравнивались с координатами на географической карте. Оказалось, что ошибки в определении с геостационарной орбиты координат опорных точек (тех мысов, которые не закрыты облаками) составляли 1,8–2,6 км по поверхности Земли.

КОНТРОЛЬНЫЕ ЗАДАНИЯ

1. Перечислите основные методы распознавания изображений. Каковы их особенности?
2. Как выбрать оптимальный порог при пороговой сегментации?
3. Назовите этапы кластеризации при использовании алгоритма ISODATA.

4. Как производится сегментация путем наращивания областей, что такое 4-связность? 8-связность?
5. Какие способы выделения границ объектов используются в методе сегментации путем выделения границ?
6. В чем суть корреляционно-экстремального алгоритма?
7. Сформулируйте теорему Байеса. Что такое апостериорная вероятность?
8. Как найти вероятность ложной тревоги и вероятность пропуска обнаружения?
9. По каким формулам вычисляются среднее значение, дисперсия и коэффициент корреляции выборочных значений?
10. Сформулируйте, в чем состоит метод классификации по минимальным расстояниям, метод параллелепипедов, метод максимумов многомерных гистограмм.
11. В каких случаях и как следует использовать метод главных компонент?
12. В чем принципиальное отличие подхода к распознаванию изображений с использованием нейрокомпьютера от других рассмотренных в этой главе методов?
13. Проведите классификацию изображений с помощью пакета IDRISI.

ЛИТЕРАТУРА

- Яншин В.В. Анализ и обработка изображений: принципы и алгоритмы. М.: Машиностроение, 1995. 112 с.
- Валентюк А.Н., Предко К.Г. Оптическое изображение при дистанционном зондировании. Минск: Наука и техника, 1991. 360 с.
- Автоматизированная обработка изображений природных комплексов Сибири / Под ред. А.Л. Яншина и В.А. Соловьева. Новосибирск: Наука, 1988. 234 с.
- Пуятин Е.П., Аверин С.И. Обработка изображений в робототехнике. М.: Машиностроение, 1990. 320 с.
- Дистанционное зондирование: количественный подход: Пер. с англ. / Под ред. А.С. Алексеева. М.: Недра, 1983. 410 с.
- Фомин Я.А., Тарловский Г.Р. Статистическая теория распознавания образов. М.: Радио и связь, 1986. 264 с.
- Прэтт У. Цифровая обработка изображений: В 2 кн. М.: Мир, 1982. 670 с.
- Навилдис Т. Алгоритмы машинной графики и обработки изображений. М.: Радио и связь, 1986. 400 с.
- Richards J. Remote Sensing. Digital Image Analysis. An Introduction. Second Edition. Berlin: Springer-Verlag. 1993. 340 p.
- Белоглазов Н.Н., Тарасенко В.П. Корреляционно-экстремальные системы. М.: Сов. радио, 1974. 392 с.
- Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. 276 с.
- <http://code935.gsfc.nasa.gov/Tutorial>

5. ПРИМЕНЕНИЕ КОСМИЧЕСКИХ СРЕДСТВ КОНТРОЛЯ ОКРУЖАЮЩЕЙ СРЕДЫ

5.1. ГЕОГРАФИЧЕСКИЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Для анализа данных дистанционного зондирования наиболее удобны географические информационные системы (ГИС), позволяющие эффективно работать с пространственно-распределенной информацией (картами, планами, аэрокосмическими изображениями, схемами в сочетании с текстом, таблицами и др.) [1]. С данными такого рода приходится иметь дело практически в любой сфере деятельности. Это может быть карта природных ресурсов, результаты экологического мониторинга территории, атлас земельного кадастра, план городских кварталов, схема движения транспорта и др. ГИС позволяет накапливать, интегрировать и анализировать информацию, оперативно находить нужные сведения и отображать их в удобной для использования форме, оценивать геометрические характеристики объектов (площадь лесного массива, длину улицы, расстояние между городами).

Геоинформационная система мониторинга окружающей среды обычно включает блоки:

- сбора информации об объектах окружающей среды;
- обработки, сортировки, запоминания и хранения информации;
- моделирования и выявления взаимосвязей между процессами;
- оценки текущего состояния окружающей среды;
- прогнозной оценки (расчета тренда) состояния окружающей среды;
- выполнения специальной обработки данных по заданию пользователя.

5.1.1. ГИС мониторинга лесных пожаров*

Эта ГИС предназначена для информационной поддержки процессов принятия решений в системе охраны леса и управления лесными пожарами, снижения наносимого ими ущерба природе и обществу, расширения имеющихся данных о процессах возникновения и развития лесных пожаров, их влиянии на структуру и динамику лесного фонда.

* При написании п. 5.1.1 использованы материалы отчета о совместном проекте «Разработать геоинформационную систему и методы использования космической информации для мониторинга лесных пожаров» Международного института леса (Москва) и Института леса СО РАН (Красноярск), научный руководитель — академик А.С. Исаев.

и также для обеспечения всем пользователям ГИС доступа к информации о лесных пожарах и их воздействии на лесные экосистемы [2].

Мониторинг лесных пожаров обеспечивает решение комплекса функциональных задач, связанных с оценкой и прогнозом пожарной опасности в лесах, обнаружением загораний в лесу и лесных пожаров, контролем их состояния и динамики, оценкой последствий лесных пожаров, управлением работой лесопожарных служб. Источниками информации для решения таких задач являются наземные и авиационные службы охраны лесов от пожаров, из них особо значимы космические средства дистанционного зондирования [2, 3].

Космический сегмент охраны лесов от пожаров формируется на основе привлечения спутниковых средств наблюдения общего назначения (метеорологические спутники, спутники для исследования природных ресурсов и др.), а также создания специализированных космических систем (КС) наблюдения за лесопожарной обстановкой и (или) размещения специальных приборов наблюдения на борту космических аппаратов общего назначения. При этом лесные пожары обнаруживаются на территории лесного фонда России, ограниченной 40–72° с. ш. и 20–170° в. д., общей площадью 1180 млн га, а особое внимание привлекает территория в пределах 50–70° с. ш. и 30–170° в. д.

На федеральном, региональном и локальном уровнях ГИС мониторинга лесных пожаров включает пять функциональных подсистем:

- оценки и прогноза пожарной опасности в лесах;
- мониторинга процессов возникновения и развития лесных пожаров;
- мониторинга процессов предупреждения, обнаружения и тушения лесных пожаров;
- интеллектуальной поддержки управленческих решений;
- оценки последствий лесных пожаров и результатов функционирования системы охраны леса.

Остановимся подробнее на первых двух подсистемах.

Подсистема оценки и прогноза пожарной опасности в лесах. Предназначена для оценки угрозы возникновения и развития лесных пожаров, а также определения условий функционирования подсистемы охраны леса. На основе получаемых оценок и прогнозов осуществляется: регулирование уровня хозяйственной деятельности в лесах и ограничение доступа к лесным массивам; назначение режимов воздушного и наземного патрулирования лесов, дежурства на пожарных наблюдательных пунктах и станциях; регулирование степени готовности сил и средств пожаротушения, количества вводимых в действие ресурсов; информирование общественности об угрозе возникновения и развития лесных пожаров.

Комплекс задач, решаемых в рамках данной подсистемы, включает в себя: долго-, средне- и краткосрочное прогнозирование пожарной опасности в лесах по условиям погоды; определение границ снежного покрова и сроков наступления пожароопасного сезона; определение границ грозовой облачности и интенсивности молниевых разрядов; определение текущего влагосодержания лесных горючих материалов (ЛГМ); оценку напряженности пожароопасных периодов и сезонов.

Долгосрочное прогнозирование пожарной опасности в лесах выполняется на основе долгосрочных (сезонных, месячных) прогнозов погоды с использованием прогностических зависимостей нарастания показателей пожарной опасности от числа и длительности засушливых периодов, температуры воздуха и количества атмосферных осадков. В качестве основных показателей пожарной опасности используют комплексный метеорологический показатель и показатели влагосодержания лесных горючих материалов.

Целью долгосрочного прогнозирования является определение зон и периодов (сезонов) высокой и чрезвычайной пожарной опасности, требующих усиления профилактических противопожарных мероприятий, заблаговременного наращивания сил и средств пожаротушения, приведения их в состояние повышенной готовности. Влагосодержание ЛГМ в зонах высокой и чрезвычайной пожарной опасности оценивается уравнением водного баланса, приходную часть которого составляют атмосферные осадки, подземный и поверхностный приток воды, поглощенная из воздуха влага, а расходную часть — физическое и физиологическое испарение, поверхностный и подземный стоки.

Организационные уровни решения задачи долгосрочного прогнозирования пожарной опасности — федеральный и региональный.

Среднесрочное прогнозирование пожарной опасности выполняется на основании среднесрочных (3–10 суток) прогнозов погоды, прогнозов аномалий температуры воздуха, предупреждений об особо опасных гидрометеорологических явлениях, специализированных прогнозов развития конвективной облачности. В качестве основных показателей пожарной опасности, предупреждающих об угрозе возникновения и развития лесных пожаров, наряду с комплексным метеорологическим показателем и показателями влагосодержания ЛГМ используются интегральные характеристики интенсивности пожаров — скорость распространения огня и интенсивности тепловыделения. Эти параметры определяются структурными и теплофизическими характеристиками ЛГМ, их влагосодержанием, рельефом местности и скоростью ветра. Приближенный характер прогноза основных метеоэлементов, значительное варьирование структурных и теплофизических характеристик

ЛГМ оправдывают использование для их оценки упрощенных моделей динамики лесных пожаров.

Среднесрочное прогнозирование позволяет по условиям погоды строить прогностические карты пожарной опасности в лесах, необходимые для текущего планирования работы лесопожарных служб и мероприятий по профилактике, обнаружению и тушению лесных пожаров, а также своевременно маневрировать силами и средствами пожаротушения.

Организационные уровни решения задачи среднесрочного прогнозирования пожарной опасности — региональный и федеральный.

Краткосрочное прогнозирование и оценка текущей пожарной опасности выполняются на основе краткосрочных прогнозов погоды, данных метеонаблюдений, данных дистанционного зондирования и показателей пожарной опасности за предшествующие сутки. В качестве основных показателей пожарной опасности наряду с характеристиками влагосодержания ЛГМ и потенциальной интенсивности лесных пожаров используются вероятность возникновения и ожидаемое число пожаров.

Суточное число пожаров на территории каждого структурного подразделения системы охраны леса рассматривается как дискретная случайная величина с известной функцией распределения, параметры которой определяются влагосодержанием ЛГМ и уровнем активности основных видов источников огня. Для построения распределений и оценки их параметров используется централизованный фонд архивных данных о лесных пожарах, входящий в состав информационного обеспечения ГИС.

Краткосрочное прогнозирование и оценка пожарной опасности позволяют строить оперативные прогнозные карты, характеризующие пожарную обстановку на территории каждого структурного подразделения авиационной и наземной охраны. С учетом этой обстановки выбираются режимы и графики (маршруты) воздушного и наземного патрулирования, режимы дежурства на пожарных наблюдательных пунктах и станциях, определяются потребности в ресурсах для оперативного обслуживания возникающих лесных пожаров, общественность оповещается об угрозе возникновения и развития лесных пожаров, формируются заявки на воздушные суда и мобилизацию дополнительно привлекаемых ресурсов.

Организационные уровни решения задачи краткосрочного прогнозирования и оценки текущей пожарной опасности — региональный и локальный.

Для определения границ снежного покрова и сроков наступления по-
жароопасного сезона используют дешифрирование изображений мас-

штаба 1:1 000 000 и крупнее, ежедневно получаемых со спутников по всей территории лесного фонда страны, а также обработку данных радиометрических съемок в диапазоне 0,8–10 см и прогнозов погоды.

Знание границы снежного покрова позволяет определить сроки наступления пожароопасного сезона, что важно для прогнозирования сроков авиационной и наземной охраны лесов, контролируемых выжиганий растительности, применения сил и средств пожаротушения в широтном направлении по мере продвижения пожароопасного сезона с юга на север. За начало пожароопасного сезона принимается дата устойчивого схода снежного покрова — полного освобождения открытых лесных участков от снега. Наличие существенной задержки схода снежного покрова на покрытых лесом площадях по отношению к открытым участкам позволяет обеспечить необходимую заблаговременность в определении сроков проведения противопожарных мероприятий на активно охраняемой территории лесного фонда.

Организационные уровни решения задачи определения границ снежного покрова и сроков наступления пожароопасного сезона — федеральный и региональный.

Для *определения границ грозовой облачности* используют дешифрирование изображений, получаемых с искусственных спутников Земли, и обработку данных многоканальных радиометрических съемок. Для этого перспективны видимый (0,4–0,7 мкм) и ближний инфракрасный (0,7–1,0 мкм) диапазоны. Индикатором развития гроз считается кучево-дождевая облачность, которая имеет большую вертикальную мощность и значительные (до нескольких километров) поперечные размеры и потому хорошо видна на космических снимках. Вследствие низкой температуры верхней границы грозовых кучево-дождевых облаков они просматриваются и на ИК-снимках.

По космическим изображениям невозможно определить признаки «сухих» гроз, развивающихся на верхних и вторичных атмосферных фронтах, а также при внутримассовых грозах, которые наблюдаются в антициклонах, в размытых барических полях и реже — в теплом секторе циклона. Однако применение программного пакета восстановления вертикального профиля температуры и влажности воздуха должно облегчить решение этой задачи.

Исключительно важная роль сухих гроз в появлении массовых вспышек лесных пожаров обуславливает необходимость использования для их дешифрирования данных наземных метеорологических и радиолокационных измерений. При этом наиболее достоверными источниками информации об интенсивности молниевых разрядов оказываются специализированные системы грозопеленгации путем приема радиосигналов от разрядов молний. Дешифрирование крупномасштаб-

ных метеорологических явлений, выделение районов активной грозовой деятельности и прослеживание их перемещения по территории позволяют прогнозировать процессы возникновения лесных пожаров от природных источников огня и прежде всего — массовых вспышек лесных пожаров, приводящих к пиковым нагрузкам в системе охраны леса и выходу пожаров из-под контроля этой системы.

Организационный уровень решения задачи определения границ грозовой облачности и интенсивности молниевых разрядов — региональный и локальный.

Для *определения границ ресурсной облачности* и для искусственного вызывания осадков используют дешифрирование снимков с космических аппаратов, а также обработку данных радиометрических съемок. Для этого также перспективны видимый и ближний ИК-диапазоны. Время съемки с 11 до 15 часов местного времени. Для искусственного вызывания осадков наиболее приемлемыми считаются мощные конвективные облака, хорошо просматриваемые на снимках из космоса в виде светлых зернистых пятен различного диаметра. Изображение ресурсной облачности на снимках отличается большой яркостной неоднородностью. Тон изображения может меняться от серого до ярко-белого, причем светлый тон изображения чередуется с темным, соответствующим изображению земной поверхности, видимой между облаками.

Дешифрирование и картирование ресурсной облачности, а также прослеживание ее динамики во времени и по территории позволяют определять интенсивность развития конвекции и выделять зоны оперативной работы по искусственному вызыванию осадков. Такие дешифрирование и картирование производятся в районах действия крупных лесных пожаров. На космических снимках в районе пожара выделяются две зоны мощной кучевой облачности. Первая из них радиусом 30—50 км от центра пожара перспективна для оперативной работы по вызыванию осадков в ближайшие часы, вторая радиусом 100—150 км — фоновая и определяет интенсивность развития конвекции в районе пожара.

Отсутствие надежных методов оценки количества конвективных облаков, из которых можно вызывать осадки, обуславливает необходимость дополнительного привлечения данных радиолокации и радиозондирования. Оперативное и обоснованное принятие решений о тушении крупных лесных пожаров искусственно вызываемыми осадками — одно из важных направлений повышения эффективности работы авиалеоохраны.

Организационный уровень решения задачи определения границ ресурсной облачности для искусственного вызывания осадков — региональный и локальный.

Для *определения текущего влагосодержания ЛГМ* используют космические сканерные съемки лесной территории в ИК- и РТ-диапазонах и последующую обработку изображений. Регистрация теплового излучения территории производится в оптическом диапазоне 8–14 мкм (максимум излучения природных объектов) и в радиодиапазоне 0,3–3,3 см с учетом «окон» прозрачности атмосферы.

При обработке производится пространственное совмещение изображений, вычисление радиационной и радиояркостной температур, аккумулированных за бездождевой период, а также разностей каждой из этих температур между смежными циклами съемки и последующая приближенная оценка влагосодержания ЛГМ. Эти вычисления базируются на уравнениях связи между спектральными коэффициентами яркости, тепловой инерцией и влагосодержанием ЛГМ.

Определение текущего влагосодержания ЛГМ позволяет корректировать краткосрочные прогнозы пожарной опасности и диагностировать фактическую пожарную обстановку на территории лесного фонда, а соответственно и корректировать управленческие решения в системе охраны леса.

Организационный уровень решения задачи оценки текущего влагосодержания лесных горючих материалов — региональный и локальный.

Оценка напряженности пожароопасных периодов и сезонов основана на данных краткосрочных прогнозов и результатов диагностики фактической пожарной опасности, а также на данных учета горимости лесов в конкретном пожароопасном периоде и сезоне. Относительная напряженность пожароопасного периода обусловлена потенциальной длительностью распространения возникших пожаров и потенциальным объемом работ по борьбе с огнем с учетом скорости его распространения и интенсивности тепловыделения. За условную оценку, независимую от фактического числа загораний, принимается общее расстояние пройденное огнем от единичного постоянно действующего источника.

Общая напряженность пожароопасного сезона определяется как сумма напряженностей входящих в него пожароопасных периодов. Используя эти параметры, можно получить интегральные характеристики условий работы структурных подразделений системы охраны леса как основу объективной оценки результатов и эффективности их функционирования.

Организационные уровни решения задачи оценки напряженности пожароопасных периодов и сезонов — федеральный, региональный и локальный. Для локальных систем суммируются напряженности для всех входящих в них структурных подразделений, а для региональных и федеральных систем охраны леса — напряженности соответственно для входящих в них локальных и региональных систем.

Подсистема мониторинга процессов возникновения и развития лесных пожаров. Данная подсистема предназначена для регистрации возникающих и действующих лесных пожаров, оценки их параметров, формирования и ведения оперативных и архивных баз данных о лесных пожарах, слежения за состоянием и динамикой горимости лесов, формирования ежедневных, месячных и сезонных карт, сводок и отчетности о лесных пожарах, оперативного и ретроспективного анализа горимости лесов, анализа временной и пространственной структур горимости лесов, а также источников огня и причин возникновения лесных пожаров. На основе получаемой информации о процессах возникновения и развития лесных пожаров принимаются основные управленческие решения по организации их тушения и ликвидации последствий, а также обеспечивается информационная поддержка процессов проектирования и планирования противопожарных мероприятий на федеральном, региональном и локальном уровнях.

Организационные уровни решения задачи регистрации лесных пожаров и ведения оперативных баз данных — локальный, региональный и федеральный. База оперативных данных о лесных пожарах ГИС локального и регионального уровней содержит информацию о каждом пожаре, обнаруженном на контролируемой ею территории, а также интегральные характеристики горимости лесов за каждый день пожароопасного сезона в разрезе структурных подразделений системы охраны леса.

Базы оперативных данных федеральной ГИС содержит информацию о каждом крупном лесном пожаре на территории страны, а также интегральные характеристики горимости лесов за каждый день пожароопасного сезона в разрезе областей (краев), республик в составе Российской Федерации и региональных авиабаз.

Оценка параметров лесных пожаров и контроль за их состоянием выполняется в результате пространственной регистрации (картирования) крупных лесных пожаров с определением контуров пожара и размеров пройденной огнем площади, энергетической диагностики кромки лесных пожаров с определением интенсивности тепловыделения, обработки данных о дымовых шлейфах в конвенционных колонках, образующихся при крупных лесных пожарах, а также сопоставления полученных при этом данных с результатами визуальных наблюдений за крупными лесными пожарами. Получаемая при этом информация необходима для оценки состояния и прогноза динамики крупных лесных пожаров, а также потребности в силах и средствах для их тушения, планирования работ по борьбе с огнем и оценки воздействия крупных лесных пожаров на окружающую природную среду. В рамках этой задачи производятся контроль числа и площади крупных лесных пожа-

ров, оценка задымленности территории, а также обнаружение скрытых очагов горения и лесных пожаров, возникающих в условиях задымленности.

Космические методы контроля числа и площади крупных лесных пожаров, на долю которых приходится свыше 80% всей охватываемой огнем площади лесного фонда, обеспечивают существенное повышение точности учета горимости лесов и объективности оценки последствий лесных пожаров, наносимого ими ущерба природе и обществу.

Организационные уровни решения задачи оценки параметров и контроля за их состоянием — федеральный, региональный и локальный.

Информационное обеспечение ГИС мониторинга лесных пожаров

Информационные потоки для обеспечения ГИС лесных пожаров, с точки зрения периодичности обновления данных, подразделяются на условно-постоянные, сезонные и оперативные.

К *условно-постоянным* относятся данные, обновляемые не чаще одного раза в сезон:

- информация административного характера (справочники по структурным подразделениям системы охраны лесов, по метеорологическим станциям);
- характеристики лесного фонда, не изменяющиеся в течение пожароопасного сезона;
- параметры моделей, используемых при решении функциональных задач.

К *сезонным* относятся данные, изменяющиеся несколько раз в течение пожароопасного сезона: сведения о фенологическом состоянии растительности; долгосрочные прогнозы погоды.

К *оперативным* относятся данные, изменяющиеся с периодичностью в сутки и чаще:

- о текущих метеонаблюдениях;
- о структуре и степени активности источников огня;
- текущие оценки и прогнозы пожарной опасности.

Основной объем составляют метеоданные, космические дистанционные данные и данные о возникающих и действующих лесных пожарах. После регистрации в специальных базах эта информация служит основой решения большинства задач мониторинга лесных пожаров на всех уровнях.

Дистанционные данные используются для определения границ снежного покрова и сроков наступления пожароопасного сезона, выделения районов грозовой и ресурсной облачности, оценки текущего влагосодержания ЛГМ и степени текущей пожарной опасности, регистрации возникающих и действующих лесных пожаров, слежения за пе-

развитием и интенсивностью, определения послепожарной структуры лесных земель и степени повреждения насаждений, оценки динамики гарей и поврежденных пожарами насаждений.

Общий объем требуемых каждым уровнем ГИС дистанционных данных можно оценить по площади обслуживаемого конкретной ГИС лесного фонда и ориентировочным показателям необходимой периодичности их поступления. Для регионального уровня ГИС, например, выдвигаются следующие требования:

- обзор территории для обнаружения и слежения за динамикой пожаров в светлое время суток — каждый час, в ночное время — каждые три часа;

- получение информации о шлейфах и конвекционных колонках — два раза в сутки;

- получение данных об облачности — три раза в сутки;

- уточнение границ классов пожарной опасности — один раз в сутки;

- получение данных о текущем влагосодержании лесных горючих материалов в период высокой пожарной опасности — каждые четыре часа;

- уточнение границ схода снежного покрова — каждые трое суток.

Поступающая на вход ГИС любого уровня оперативная информация о действующих и крупных очагах пожаров содержит:

- время обнаружения пожаров;

- координаты очагов;

- границы, размеры и интенсивность тепловыделения кромки огня;

- направление распространения огня;

- характеристики дымовых шлейфов и конвекционных колонок крупных пожаров;

- сведения об опасных явлениях в зонах пожаров;

- данные о районах активности грозовой деятельности с прогнозом развития их динамики;

- данные о районах конвективной облачности;

- данные о задымленности территории в зонах пожаров.

Информация о крупном лесном пожаре фиксируется в карточке крупного лесного пожара, где указываются динамика его развития и организация тушения, и в отчете о ходе его ликвидации. Интегральные характеристики горимости лесов отображаются в ежедневном сводном отчете о действующих лесных пожарах, содержащем сведения о числе пожаров на день отчета (в том числе возникших за день и нарастающим итогом с начала пожароопасного сезона), пройденной ими площади с разделением ее на лесную и нелесную (и выделением площади, пройденной верховыми пожарами), числе пожаров, ликвидированных в день

обнаружения, числе работающих на тушении людей и технических средств.

К основным видам выходной картографической документации относятся схематические карты противопожарных мероприятий, справочные лесные карты и т.д. Выходные картографические документы представляются в следующих масштабах:

- от 1:10 000 000 до 1:2 500 000 на федеральном уровне;
- от 1:2 500 000 до 1:500 000 на региональном уровне;
- от 1:500 000 до 1:100 000 на локальном уровне;
- от 1:100 000 до 1:25 000 на детальном уровне.

В зависимости от способа воспроизведения они разделяются на три группы:

- визуализируемые на экране;
- выдаваемые в растровом виде на принтерах;
- вычерчиваемые в векторном виде на графопостроителях.

По содержанию и оформлению выходные картографические документы должны удовлетворять требованиям государственных и отраслевых стандартов. Детально должны показываться элементы топографической основы: железные и автомобильные дороги, гидрографическая сеть, населенные пункты и различные виды угодий — сенокосы, пастбища, кустарники, болота.

Организация баз данных по уровням ГИС. Информационной базой системы является единый банк пространственно-совмещенных картографических и фактографических данных о лесном фонде и происходящих в нем изменениях. Структурно этот банк может состоять из набора взаимодействующих баз данных, обменивающихся между собой информационными потоками. По форме представления хранимой информации базы подразделяются на карто- и фактографические. В первых основными элементами являются данные, представленные в векторном и растровом виде, а во вторых — количественные и качественные данные в числовом и символьном виде.

Из банков федерального, регионального и локального уровней можно получать информацию о лесном фонде разной степени генерализации. Различные уровни ГИС связаны через базы оперативных данных и последствий пожаров, причем в информации об изменениях на вышележащем уровне, например на федеральном, должны быть объединены данные из соответствующей базы нижележащего (регионального) уровня.

Базы структурированы по реляционному типу, означаящему, что вся информация в них размещается в реляционных таблицах, операциях над которыми выполняются по правилам реляционной алгебры. Это позволяет развязать программные средства и информацию, сделать их

Независимыми, что является важнейшим требованием, предъявляемым к базам данных.

Для каждого графического объекта в соответствующей таблице отводится строка, содержащая характеристики объекта и координаты его точек. Использование векторного формата позволяет экономить память и сохранять точность представления топографической основы. Изменения графической информации в зависимости от их вида вносятся не только в векторном, но и в растровом формате, как и выдача картографических материалов. Это возможно благодаря применению программ перевода информации из одного формата в другой.

Для ГИС любого уровня ключевым моментом организации данных является совмещение тематической и картографической информации, для чего тематическим данным присваивается номер того объекта, который они описывают. Это позволяет, например, изучать изображенный на экране картографический объект, указывая на него курсором и запрашивая его фактографические характеристики из совмещенной базы данных.

Накапливаемая в картографической базе информация поступает из различных источников и в различное время: из таблиц гео данных, из фотоабрисов на космических и аэрофотоснимках, с лесоустроительных картографических материалов прошлых лет, с топографических карт и пр. Иными словами, на стадии ввода информация порождается различными системами координат. Впоследствии она приводится к единой системе координат.

Для обеспечения необходимой точности планово-картографических материалов принят способ наложения тематической информации на топографическую основу, которая получается в результате геодезической съемки (гео данных) и в этом случае гарантирует геометрическую точность информации. При этом способе цифровой картографии, называемом «каркасным подходом», картографическая информация разделяется на каркас и заполнение, а система ввода графической информации имеет два канала: высокоточный, удерживающий ошибки в заданных пределах и служащий для топоосновы, и менее точный канал заполнения каркаса дистанционными данными и абрисами. При этом точность картографии обусловлена точностью каркаса (топоосновы), а ошибки «колеблются» лишь внутри ячеек этого каркаса.

Первый канал использует векторный формат представления, не имеющий ошибки дискретизации; второй — на начальном этапе растровый формат, который затем переводится в векторный.

Для хранения контурной информации в масштабе 1:25 000 в векторном формате расходуется дисковая память из расчета 200 байт на лесной выдел; для хранения картографической базы данных на лесные

кварталы с транспортной сетью, гидрографией и объектами инфраструктуры в масштабе 1:100 000 требуется порядка 2 кбайт на квартал.

Опыт создания цифровых картографических баз данных свидетельствует, что целесообразно хранить цифровые карты, различающиеся по масштабам в 4–5 раз, а карты промежуточных масштабов строить методом генерализации цифровой карты ближайшего более крупного масштаба. Исходя из этого, применяемые в ГИС мониторинга лесных пожаров картографические данные могут быть получены при наличии цифровых карт следующего масштабного ряда: 1:25 000, 1:100 000, 1:1 000 000, 1:2 500 000 и 1:10 000 000.

Структура совмещенных баз данных (рис. 5.1, где НСИ — база нормативно-справочной информации, ТЭП — базы технико-экономических показателей) определяет потоки информации, действующие внутри ГИС каждого уровня при решении различных задач. Главное отличие таких структур состоит в степени объединения и генерализации используемой информации на разных уровнях ГИС. При этом база фенологических данных (БФД) содержит даты установления и схода снежного покрова, характеристики текущего влагосодержания подстилающей поверхности и лесных горючих материалов, лесопирологические показатели, классы пожарной опасности, карты пожароопасных периодов.

База метеорологических данных (БМД) включает долго-, среднесрочные и краткосрочный прогнозы пожарной опасности в лесах по условиям

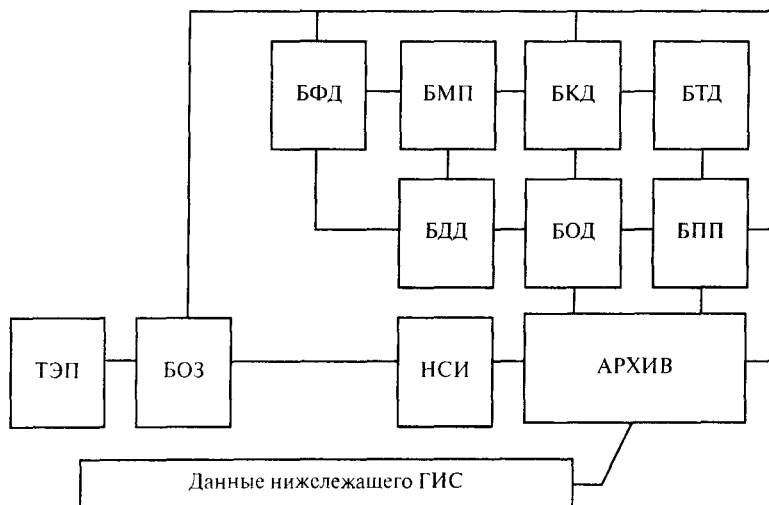


Рис. 5.1. Общая структура совмещенных баз данных

погоды, класс пожарной опасности по условиям погоды, температуру воздуха и почвы, их влажность, скорость и направление ветра, атмосферное давление, комплексный показатель горимости лесов, количество проникающих под полог леса осадков в зависимости от возраста, состава, густоты древостоя и развития подлеска, границы грозовой и ресурсной облачности, интенсивность молниевых разрядов и, возможно, некоторые другие показатели.

База охраны и защиты леса (БОЗ) объединяет сведения о размещении средств и ресурсов системы охраны леса (авиационных и наземных пожарных команд, воздушных судов и наземных транспортных средств, пожарной техники и оборудования), степени их готовности, природно-экономических условиях охраняемой территории и т.д. Данные о выполненных работах по профилактике, обнаружению и тушению лесных пожаров хранятся в БОЗ локального и регионального уровней, а обобщенная информация о деятельности региональных систем охраны леса и учетные данные о работе лесопожарных сил быстрого реагирования — в БОЗ федерального уровня.

В базе картографических данных (БКД) хранятся цифровые карты различных масштабов с точечно-локализованными объектами инфраструктуры, узлами и линиями картографической сетки, рельефом, орографической сетью с горными системами, гидрологической и дорожной сетью, границами регионального, областного и районного деления, границами предприятий, лесных массивов, зон действия авиаотделений и другой необходимой картографической информацией, например физико-географическим районированием.

База тематических данных (БТД) объединяет лесоустроительную информацию в соответствии с уровнем ГИС.

База оперативных данных (БОД) содержит дату обнаружения пожара, источник информации о пожаре, тип пожара (низовой или верховой) и его интенсивность, имя региона и области, географические координаты очага и опорных точек, условные координаты очага и опорных точек в системе координат картографической базы данных, степень задымленности территории, контур границы зоны лесного пожара, дату локализации пожара, дату окончания пожара, дату выпадения способных потушить пожар осадков, контур и площадь пожарища и, возможно, некоторые другие данные.

База последствий пожаров (БПП) включает название региона, области и предприятия, географические и условные координаты опорных точек, контур границы пожарища, площадь пожарища и состояние, категорию лесов, категорию земель и преобладающую породу на пройденном пожаром участке, степень повреждения насаждений по породам, прогноз состояния насаждений на ближайшие 2—3 года, вероят-

ность массового размножения насекомых-вредителей как возможности угрозы соседним насаждениям, оценку прямого и косвенного ущерба, размеры ущерба по преобладающей породе, а также среднего и предотвращенного ущерба, оценку размеров сохраненной от огня площади лесного фонда, некоторые другие данные, определяемые по материалам лесоустройства и данным наземных обследований. Эта база объединяет данные о последствиях пожаров, полученные из нижележащих уровней ГИС.

В периодически обновляемую базу дистанционных данных (БДД) внесены цифровая запись необработанного или предварительно обработанного изображения (обработанное изображение может быть картографическим файлом), тип съемочной аппаратуры, вид носителя исходного изображения (негатив, позитив, запись на компакт-диске или на цифровой магнитной ленте), номер спектрального канала или спектральный диапазон, дата и время съемки, высота Солнца в момент съемки, высота орбиты спутника, характер витка (восходящий или нисходящий), число элементов в строке, число строк, число бит/пиксел, разрешение на местности, географические координаты снимка и центра изображения, данные об опорных и контрольных точках, орбитальные или навигационные параметры, параметры состояния бортовых систем, корректирующие коэффициенты и, возможно, некоторые другие данные, например параметры земного эллипсоида и т.д.

База архивных данных (Архив) содержит информацию о пожароопасном сезоне, коды области, края, авиаотделения, лесхоза, лесничества, предприятия, данные о принадлежности лесов, координаты и номер пожара, время и способ его обнаружения, скорость ветра, причину пожара, пройденную огнем площадь с разделением ее на лесную и не лесную и с выделением площади, пройденной верховым и подземным пожаром, категории лесов и земель, преобладающую породу, напочвенный покров, способы и средства тушения пожара, расстояния (до населенного пункта, транспортных путей и места высадки пожарного десанта), затраты авиабазы, стоимость работ по тушению, общий ущерб, время выпадения фронтальных осадков и т.д.

Элементы ГИС мониторинга лесных пожаров внедрены в Сибирском региональном центре МЧС на федеральном уровне и в Красноярской базе авиационной охраны лесов — на региональном. Используется пакет прикладных программ ArcView.

5.2.1. Общая характеристика лесного пожара

Типовую схему возникновения и развития лесного пожара (ЛП) можно представить следующим образом [2, 3]. На территории, например, северной зоны тайги в Красноярском крае (сосняки лишайниковые) действует устойчивый антициклон, периодически пересекаемый холодными атмосферными фронтами, которые сопровождаются грозными явлениями и сильными ветрами. Количество жидких осадков в этих случаях, как правило, невелико (3–5 мм). Такая ситуация часто ведет к возникновению лесных пожаров от гроз, обычно в непредсказуемых и удаленных от населенных пунктов местах. В центральной и южной зоне Красноярского края большинство пожаров имеют антропогенное происхождение.

Возникший пожар имеет характерную форму кольца, вытянутого в направлении ветра. При этом подветренная часть кромки (фронт) имеет увеличенную ширину по сравнению с наветренной (тыл), причем их отношение пропорционально скорости ветра под пологом древостоя. Так, в сосняке лишайниковом при скорости ветра 3 м/с (4-й класс пожарной опасности) и температуре воздуха 300 К через 25–30 мин сформируется кольцо пожара, занимающее площадь около 0,4 га. Скорость фронта составляет около 2 м/мин, тыла — 0,3 м/мин. Ширина фронтальной кромки будет примерно равной 2,0 м, в том числе пламенной зоны 1,2 м и зоны горения углеродистых остатков 0,8 м. Для тыльной кромки эти значения составляют 0,3 и 0,1 м соответственно. Излучательные характеристики представляются в виде:

- на фронте средняя термодинамическая температура пламенной зоны $T_n = 1100$ К, углеродистой зоны $T_y = 800$ К, пульсации температуры в газовой фазе $\Delta T = 200$ К, средняя частота пульсаций $f = 10$ Гц;

- в тыльной области температуры $T_n = 1000$ К и $T_y = 800$ К.

В соответствии с формулой Планка при $T = 1100$ К максимум излучения приходится на $\lambda = 2,6$ мкм, при $T = 1000$ К — на $\lambda = 2,9$ мкм, при $T = 800$ К — на $\lambda = 3,6$ мкм. Коэффициент теплового излучения пламени $\epsilon = 0,8$. Спектральные особенности наблюдаются на полосах поглощения воды и углекислого газа. Для углеродистой зоны коэффициент излучения близок к 1. Эти данные позволяют оценить средний по кромке ЛП лучистый поток в диапазоне 2–14 мкм. Он составляет около 80 кВт/м². Мощность излучения кромки горения площадью 300 м² равна 24 000 кВт, т.е. лишь около 20% энергии пожара, остальные 80% расходуются на прогрев слоя атмосферы над пожаром (конвекционной колонки), почвы и несгоревших материалов.

Восходящее от пожара излучение ослабляется пологом древостоя и атмосферой. Коэффициент пропускания ИК-излучения пологом варьируется от 0,3 до 0,9. Принимая его среднее значение 0,6, получаем мощность результирующего лучистого потока, выходящего в атмосферу, $\Phi = 14\,000$ кВт.

Леса Сибири почти ежегодно подвергаются массовым пожарам (в среднем около 30 тыс.). За 2–3 месяца сгорает около $2 \cdot 10^7$ т биомассы, в атмосферу выбрасывается около $2 \cdot 10^6$ т продуктов горения, дымовые шлейфы тянутся на десятки километров, эмиссия ЛП поднимается на 5–7 км. Даже в районах, непосредственно не подвергнутых техногенному влиянию, в эмиссиях обнаруживаются тяжелые металлы и другие элементы с концентрацией, сильно отличающейся от фоновой. В целом по земному шару выбросы в атмосферу от ЛП приблизительно равны выбросам от вулканов.

При обычных метеоусловиях пожары не причиняют существенно го вреда и даже в ряде случаев приносят пользу лесному хозяйству. Они очищают лес от уже погибших деревьев и кустарников, тем самым снижая пожарную опасность, уничтожают очаги насекомых-вредителей и т.д. Такие пожары способствуют возобновлению леса. На севере Сибири от 5 до 15% семян деревьев некоторых хвойных пород сохраняются в шишках в течение 3–4 лет, термическое воздействие пожара помогает шишкам раскрыться и обеспечивает обсеменение площади гари.

Вместе с тем на фоне длительной засухи пожары действуют как разрушители природной среды, особенно в зонах вечной мерзлоты.

5.2.2. Космический лесопожарный мониторинг

Северные (бореальные) леса России занимают площадь около 600 млн га, почти четверть из них приходится на Красноярский край. Зона лесов характеризуется малой плотностью населения. В настоящее время спутниковые данные — это единственно доступная информация о лесных пожарах в неохраняемых северных лесах, существенно дополняющая данные, получаемые с помощью обычных методов на охраняемых территориях. Космический лесопожарный мониторинг способен решать следующие задачи [4–6]:

- оценивать пожарную опасность лесных регионов и прогнозировать ее динамику;

- обнаруживать ЛП и определять их интенсивность, а также прогнозировать поведение крупных пожаров;

- проводить инвентаризацию послепожарного состояния лесов.

Пожарная опасность лесных регионов и возможность возникновения ЛП определяются влагосодержанием проводников горения [4–6]. К их

числу относятся: напочвенный покров (при низовых пожарах), подстилка и торф (при подземных пожарах), древесно-кустарниковый ярус (при верховых пожарах). Подавляющее большинство ЛП возникает в результате загорания напочвенного покрова, что позволяет ограничиться при оценке условий их возникновения расчетом влагосодержания только этой группы лесных горючих материалов (ЛГМ).

Каждый вид ЛГМ в напочвенном покрове характеризуется предельным влагосодержанием, при котором возможно распространение горения. Для зеленых мхов оно составляет 35–40%, для кустистых лишайников — 25–35%, отмершей травы и листьев — 20–25%, лесной подстилки — 50–60%, торфа — 40–50%. Такое состояние ЛГМ обычно наступает в лесу через 85–150 часов сушки в случае отсутствия жидких осадков, при наличии солнечной радиации и слабой облачности.

В Институте леса им. В.Н. Сукачева СО РАН проведены обширные исследования сушки ЛГМ. Интенсивность высыхания лесных горючих материалов в большей мере определяется температурой материала T_m , чем температурой воздуха t , поскольку последняя не отражает напрямую процесс радиационного теплообмена ЛГМ с окружающей средой. Оказалось, что если в течение периода сушки суммировать ежедневные температуры ЛГМ, то накопленная сумма температур ΣT будет связана линейной зависимостью с влагосодержанием ЛГМ: чем выше ΣT , тем меньше влагосодержание. Коэффициенты корреляции между суммой температур ЛГМ и влагосодержанием равны $\rho = 0,68$ для травяной нетопи, $\rho = 0,88$ для опада хвои сосны, $\rho = 0,74$ для опада листвы березы.

Тесная связь между ΣT и влагосодержанием ЛГМ позволяет по сумме температур за некоторый период оценивать влагосодержание и определять пожарную опасность как степень готовности лесных горючих материалов к самопроизвольному поддержанию горения. Это позволяет использовать спутниковую информацию о радиационных температурах ЛГМ для оценки пожарной опасности. Коэффициент теплового излучения ЛГМ $\epsilon = 0,94$ – $0,96$, и потому радиационная температура T_p меньше температуры материала T_m всего на 4–6%. Максимум излучения подстилающей поверхности лежит в диапазоне 8–14 мкм, попадающем в окно прозрачности атмосферы. Таким образом, радиационная температура T_p — достаточно надежная оценка температуры ЛГМ T_m , определяемая, например, с помощью спутника NOAA (4-й и 5-й каналы сканера AVHRR). При этом учитывается, что хотя кроны деревьев и могут частично экранировать излучение, восходящее от подстилающей поверхности, но сами имеют такую же или даже несколько большую температуру, чем подстилающая поверхность.

Существующая система оценки пожарной опасности в России базируется на комплексном метеорологическом показателе Нестерова, учитывающем осадки, температуру и влажность воздуха на основании метеорологических наблюдений. Пожарная опасность ЛГМ оценивается косвенным путем по авторегрессионному соотношению

$$\Gamma_j = \xi \Gamma_{j-1} + t_j(t_j - \tau_j),$$

где τ — температура точки росы в момент наблюдений, ближайший к 12 часам местного времени, °C, t — температура воздуха в тот же момент, °C, ξ — коэффициент учета осадков, Γ_{j-1} и Γ_j — комплексные метеорологические показатели предшествующего и текущего дня.

К сожалению, очень неравномерная сеть метеостанций сокращает ее; это не позволяет надежно прогнозировать пожарную опасность по всей лесной территории. Выход из положения состоит в применении дистанционного спутникового зондирования, основанного на оценке пожарной опасности по сумме радиационных температур. В отличие от показателя Нестерова, этот метод использует информацию, получаемую непосредственно с подлежащих контролю лесных территорий.

Иначе говоря, метод состоит в суммировании поля радиационных температур подстилающей поверхности. Суммирование производится ежедневно от одного момента выпадения жидких осадков до другого по результатам съемки в 13–15 часов местного времени, когда наблюдается суточный минимум влагосодержания ЛГМ. Моменты выпадения жидких осадков определяются по данным метеостанций и информации о влажности воздуха, получаемой с радиометра HIRS спутника NOAA. Далее, используя пороговый метод классификации, участки лесной территории относят к определенному классу пожарной опасности путем сравнения с заданными порогами суммы температур для каждого пиксела. Анализ серии полей радиационных температур дает прогноз динамики пожароопасной ситуации.

Если на момент съемки часть поверхности закрыта кучево-дождевыми облаками, поля радиационных температур дополнительно анализируются на других витках, полученных в тот же день, путем взаимного наложения полей и исключения облачности.

Показатель суммы радиационных температур хорошо коррелирует с метеорологическим комплексным показателем пожарной опасности Нестерова. Так, в пожароопасные периоды 1996, 1997 гг. коэффициент корреляции составил $\rho = 0,8–0,9$.

По результатам расчета ежедневно составляются картосхемы текущей пожарной опасности (рис. 5.2), которые интегрируются в ГИС

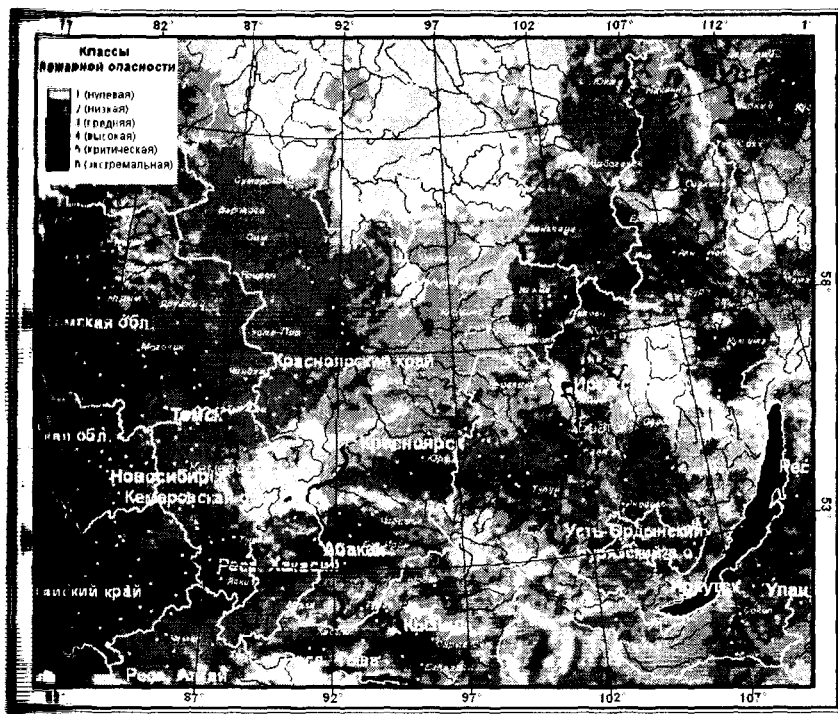


Рис. 5.2. Карта пожарной опасности в Сибири за 8 августа 2000 г. по спутниковым данным, точками отмечены населенные пункты

Мониторинга лесных пожаров, передаются в органы МЧС и авиалесоохраны. Эти сведения помогают подразделениям авиалесоохраны планировать полеты патрульных самолетов и в конечном итоге приносят экономический эффект.

Обнаружение лесных пожаров — наиболее сложная часть лесопожарного мониторинга с использованием искусственных спутников Земли [5, 6]. При обнаружении ЛП возникают трудности, связанные как с природными причинами: затененностью лесного пожара кронами, облачностью, дымом, так и с техническими: недостаточной разрешающей способностью и чувствительностью регистрирующей аппаратуры космических средств, несовершенством элементов схемы передачи и обработки информации.

В настоящее время известно несколько подходов к обнаружению ЛП с использованием данных радиометра AVHRR, установленного на спутниках серии NOAA. Вегетационные индексы, рассмотренные в п. 1.2.2, могут быть использованы не только для контроля состояния

растительности, но и для обнаружения пожаров. Нормализованный вегетационный индекс рассчитывается по формуле (1.2):

$$NDVI = (f_2 - f_1) / (f_2 + f_1),$$

где f_1 — яркость объекта в видимом диапазоне (1-й канал радиометра AVHRR — 0,58–0,68 мкм), f_2 — яркость объекта в ближнем инфракрасном диапазоне (2-й канал радиометра AVHRR — 0,725–1,0 мкм). Для растений в нормальном состоянии f_1 и f_2 заметно различаются. Интенсивность рассеянного солнечного излучения частицами дыма в этих каналах $f_1 \approx f_2$, поэтому при использовании $NDVI$ можно построить изображение, на котором шлейфы дыма от ЛП выделяются на фоне растительности (рис. 5.3). Кроме дымовых шлейфов, $NDVI$ позволяет обнаруживать оставленный сильным повреждением растительности шрам в ландшафте, который может использоваться как исходные данные для оценки площади распространения пожара.

Как уже указано, максимум теплового излучения от объекта с температурой $T = 800$ К приходится на $\lambda = 3,6$ мкм. Длина волны 3,5–4 мкм

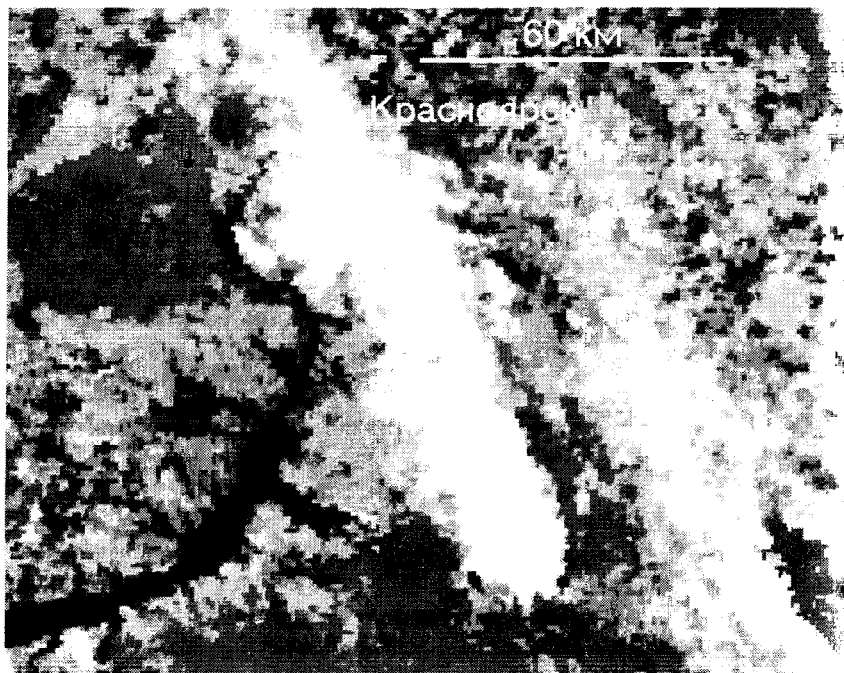


Рис. 5.3. Применение вегетационного индекса

попадает в окно прозрачности атмосферы. Сканер AVHRR спутника NOAA, 3-й канал которого работает на волнах длиной 3,55–3,93 мкм, можно эффективно использовать для обнаружения ЛП. К достоинствам спутника NOAA относятся также свободный доступ к данным, возможность наблюдать один и тот же участок поверхности 4–6 раз в сутки (при наличии двух–трех спутников), широкая (около 3000 км) полоса обзора. Однако зависимость сигнала на выходе датчика 3-го канала от интенсивности нелинейная, с выраженным насыщением.

Самым существенным недостатком сканера AVHRR является сравнительно большое мгновенное поле зрения и соответственно невысокое разрешение — $1,1 \times 1,1$ км, один пиксел вмещает изображение 121 га при наблюдении в надир. Для наиболее успешного тушения ЛП необходимо оперативно обнаруживать пожары площадью не более 1 га. Опыт авиационной охраны лесов показывает, что средняя площадь обнаруживаемых с самолета пожаров составляет 3–5 га. Такие пожары ликвидируются десантом из 4–6 человек. Площадь свыше 25 га для одного и даже двух десантов считается критической, и пожары зачастую выходят из-под контроля.

В Институте леса СО РАН в рамках геоинформационной системы мониторинга лесных пожаров разработана методика, позволяющая обнаруживать с помощью сканера AVHRR пожары площадью меньше 10 га. При малоразмерном пожаре в мгновенное поле зрения радиометра попадает как высокотемпературная зона лесного пожара, которая занимает некоторую долю p пиксела ($0 \leq p \leq 1$) и имеет температуру $T_{\text{п}}$ — 800 К, так и «холодный» фон, который занимает остальную часть пиксела ($1 - p$) и имеет температуру $T_{\text{ф}}$.

Пусть f_3 — яркость некоторого пиксела в 3-м канале. При наличии малоразмерного пожара без учета влияния атмосферы

$$f_3 = pf_{3\text{п}} + (1 - p)f_{3\text{ф}}, \quad (5.1)$$

где $f_{3\text{п}}$ — яркость от пожара, если бы он занимал весь пиксел, $f_{3\text{ф}}$ — яркость от фона, если бы он занимал весь пиксел.

По измеренному значению f_3 с использованием формулы Планка можно оценить эффективную температуру T_3 исследуемого пиксела. Наличие пожара вызывает увеличение T_3 .

На первом этапе обнаружения пожара значение T_3 сравнивается с порогом T_0 . Если $T_3 > T_0$, то возникает подозрение на наличие пожара в пределах этого пиксела. Обычно пороговое значение температуры $T_0 = 310$ –320 К. Более точный подбор T_0 обусловлен временем наблюдения и погодой. Днем в ясную солнечную погоду порог устанавливается более высоким, чем вечером или ночью.

Однако обнаружение пожаров только по значению температуры в 3-м канале затруднено наличием засветок из-за отражения солнечного излучения от облачного покрова и водной поверхности. К тому же высокое значение T_3 может быть вызвано, например, приемом излучения от нагретой солнцем поляны, полностью занимающей весь пиксел.

На втором этапе анализируются координаты пиксела. Если пиксел попадает на водохранилище, озеро, реку, то это действительно ложная засветка от водной поверхности. Для распознавания облаков используют 1-й и 2-й каналы сканера AVHRR.

На третьем этапе информация 4-го или 5-го канала позволяет отличить малоразмерный пожар от нагретой поляны или аналогичного объекта. Для пожаров с температурой $T_n = 800$ К разность температур 3-го и, например, 5-го каналов $I_n = T_3 - T_5$ будет больше такой же разности для нагретой поляны. Здесь учтено, что радиационная температура пиксела, вычисленная по яркости f_4 в 4-м или f_5 в 5-м канале, определяется в основном температурой фона $T_{\text{ф}}$, а излучение от пожара или отражение от облачного покрова не оказывают существенного влияния на f_4 и f_5 .

При наличии пожара I_n лежит в пределах 8–14 К. Величину I_n можно использовать как пожарный индекс, характеризующий интенсивность лесного пожара и зависящий от времени суток, погодных условий и времени года.

На последнем этапе уточняется площадь высокотемпературной зоны пожара. Соотношение (5.1) следует решать относительно p , так как f_{3n} можно заранее вычислить с помощью формулы излучения Планка (1.1) при стандартной T_n ; яркость фона $f_{3\text{ф}}$ можно оценить по интенсивности излучения от пикселов с низким значением T_3 . Из (5.1) находим, что относительная площадь высокотемпературной зоны

$$p = (f_{3n} - f_{3\text{ф}}) / (f_3 - f_{3\text{ф}}),$$

а абсолютная площадь $S_n = pS$, где $S = 121$ га для спутника NOAA при наблюдении в надир. Площади высокотемпературной зоны соседних пикселов суммируются.

Результаты обнаружения отображаются на карте в электронном виде (рис. 5.4), а в таблицу вносятся номер пожара, его географические координаты, субъект Российской Федерации и район, в котором он обнаружен, ближайший населенный пункт, расстояние от него до очага пожара, направление от населенного пункта на очаг пожара, площадь пожара в пикселах и гектарах. Эти сведения по каналам связи немедленно передаются в органы МЧС и авиалесоохраны.

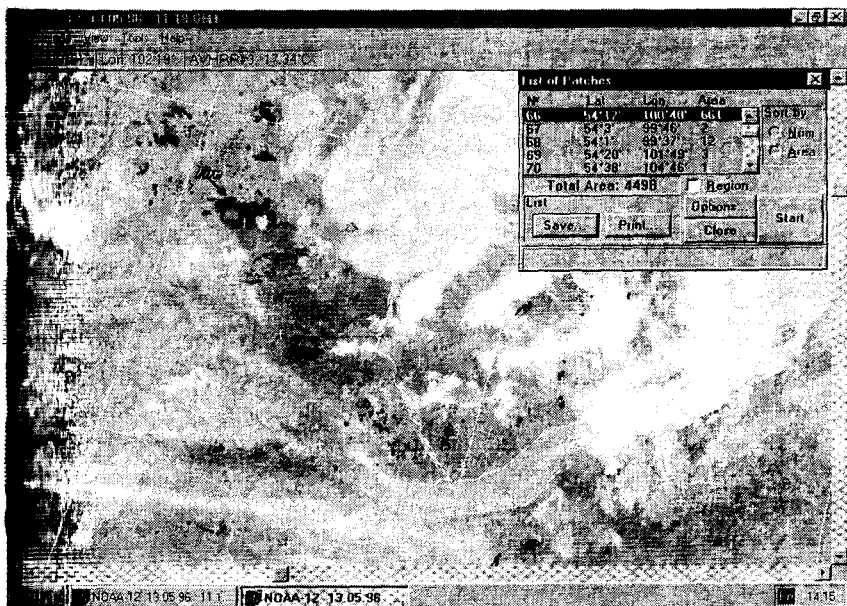


Рис. 5.4. Карта лесных пожаров, обнаруженных 13 мая 1996 г. в Северном Приангарье

Метод позволяет обнаруживать пожары полностью в автоматическом режиме. Тестирование с использованием экспериментальных пожаров показало, что с помощью сканера AVHRR можно уверенно обнаруживать пожары площадью 6 га и более.

Послепожарная инвентаризация леса, т.е. оценка последствий лесных пожаров — один из главных компонентов мониторинга лесных пожаров, — представляет значительный интерес с точки зрения лесной экологии и углеродного баланса. Эта задача с использованием космической информации решается методами распознавания образов (см. гл. 4 и [7]). Опыт показал, что при послепожарной инвентаризации лучшие результаты дает метод наращивания областей.

5.3. КОНТРОЛЬ СОСТОЯНИЯ РАСТИТЕЛЬНОСТИ

Контроль состояния наземных растительных покровов — одна из важнейших задач, решаемых методами дистанционного зондирования и включающих мониторинг посевов и пастбищ, а также лесов, лесостепных, степных и пустынных территорий. Приведем примеры использования космических средств для мониторинга сельскохозяйственных угодий и лесов [8–13].

Космический мониторинг сельскохозяйственных угодий, контроль над землепользованием, прогноз урожаев — важные составляющие современной сельскохозяйственной практики. Установлено, что биомасса, накопленная растением за время T , линейно зависит от количества солнечной радиации, поглощенной растением [8]:

$$P = \epsilon_0 S \int_0^T I(t) dt,$$

где P — продуктивность, г/(с·м²), ϵ_0 — скорость накопления биомассы, г/Дж, $I(t)$ — доля солнечной радиации, поглощаемая растением, с, S — солнечная постоянная, Вт/м². Вегетационный индекс $NDVI$ характеризует способность растения поглощать солнечную радиацию, и потому $NDVI(t)$, отнесенный к единичному интервалу времени, и доля $I(t)$ фотосинтетически активной солнечной радиации PAR приблизительно пропорциональны. Следовательно,

$$P \approx \epsilon S \int_0^T NDVI(t) dt.$$

Приведенные соотношения открывают возможность оценки по $NDVI$ прироста биомассы, урожая и других характеристик, хотя для оценки урожая требуется использовать временной ряд $NDVI$, который необходимо определять с большой точностью.

На точность определения $NDVI$ влияют многочисленные факторы [8], в том числе корректность учета зенитного угла Солнца и геометрии наблюдений, влияние атмосферы, дрейф калибровки аппаратуры, дрейф орбиты спутника, искажения изображения при наблюдении с надира и др. Установлено, например, что дрейф калибровки аппаратуры, приведший к изменению $NDVI$ всего в 0,05, приводит к погрешности в 30% при определении продуктивности пастбищ в тропиках. Разумеется, на продуктивность влияют многие факторы: сорт растения, местные условия — освещенность, температура, влажность, наличие удобрений и др., но часть из них в определенной степени отражена в $NDVI$.

Многие авторы приводят корреляционные соотношения между $NDVI$ (или другими оптическими характеристиками, измеряемыми дистанционно) и урожаем некоторых сортов сельхозкультур в определенных зонах, районах, условиях. Например, на засоленной почве урожай хлопка можно оценить как $L = -679 + 4536,8 \cdot NDVI$ (кг/га) [8]. Бо

лее простой и чаще употребительной на практике процедурой является распознавание видов сельхозкультур, растущих в некотором районе, определение их состояния, занимаемой площади и границ. Для этого используются многоспектральные данные, как правило, высокого разрешения со сканеров МСУ-Э (спутник «Ресурс-О1»), MSS (Landsat) и HVR (SPOT) [8, 9].

Наконец, спутниковая информация позволяет обнаруживать районы, пораженные засухой, вредителями, вредными техногенными выбросами. В 70–80-е годы XX в. Советский Союз закупал в больших объемах зерно за рубежом — в США, Канаде и других странах. Нет сомнения, что зарубежные партнеры при определении цены учитывали виды на урожай и использовали спутниковую информацию для оценки состояния сельхозугодий в СССР.

В последнее десятилетие в связи с появлением опасности глобальных изменений природной среды и климата возрос интерес к лесному покрову как к регулятору биосферных процессов. Леса признаются наиболее надежной природной системой, связывающей углерод и способной предотвратить возникновение парникового эффекта. В связи с этим проблема сохранения лесов и усиления их экологических функций вышла за национальные рамки и стала частью мировой экологической политики. Изучение лесов в глобальном масштабе можно реализовать только с применением искусственных спутников Земли.

Важной составляющей исследования состояния лесов является лесопожарный мониторинг, рассмотренный в § 5.3. Используя искусственные спутники Земли, удастся распознавать породный состав леса и его изменения, контролировать вырубки. Только космические средства позволяют оценивать повреждения леса на просторах Сибири, вызванные природными и антропогенными факторами.

В Институте леса СО РАН ведется наблюдение с использованием искусственных спутников Земли за состоянием уникального притундрового леса Норильской котловины (69° с. ш., 90° в. д.) на севере Красноярского края [14]. Этот лес в течение более полувека подвергается интенсивному воздействию выбросов Норильского горно-металлургического комбината (SO_2 , NO_x , тяжелые металлы). Норильская долина (см. рис. 3.1) представляет собой тектоническую депрессию длиной 150–170 км и шириной 20–60 км с многочисленными болотами и мелкими озерами. С запада она ограничена хребтом Лонтокой-Камень с максимальными высотами 600–760 м, с востока — горами Путорана (максимальные высоты, примыкающие к долине, 1000–1000 м).

Древостой образован преимущественно лиственницей, а также елью, березой, кустарниками. Ежегодный выброс SO_2 — основного за-

грязнителя воздуха, составляет около 2 млн т в год. В течение вегетационного периода растений техногенные выбросы переносятся преимущественно в юго-восточном направлении вдоль Норильской долины.

При исследованиях в 70–80-е годы применялись черно-белые фотоснимки (рис. 5.5) со спутника «Метеор-Природа» (многозональные изображения в спектральных интервалах 0,5–0,6; 0,6–0,7 и 0,7–0,8 мкм с разрешением ~50 м). Зимние снимки масштаба 1:2 500 000 применялись для дешифрирования ореола загрязнений вокруг источников эмиссий. Позднее, в 90-е годы, использовались данные, полученные с помощью аппаратуры МСУ-Э (спутник «Ресурс-О1»), MSS (спутник Landsat) и AVHRR (NOAA).

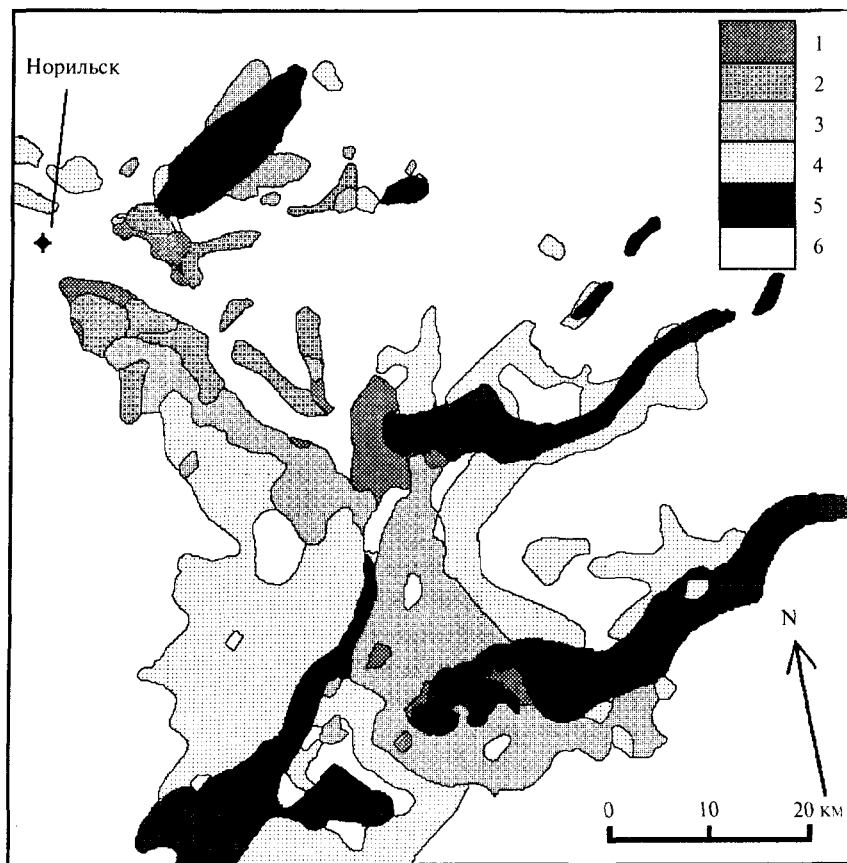


Рис. 5.5. Картограмма состояния притундровых лесов на 1979 г.:

1 — погибший древостой, 2 — сильно поврежденный, 3 — поврежденный, 4 — условно здоровый древостой, 5 — вода, 6 — нелесные территории

Фотоснимки со спутника «Метеор-Природа» вводились в ЭВМ с помощью сканера. При классификации на основе критерия Байеса по обучающим выборкам были использованы две зоны спектра (0,6–0,7 и 0,7–0,8 мкм), за классифицирующий признак принимался вегетационный индекс. При дешифрировании применялась также экспертная оценка. Точность классификации изображений и полученных картосхем определялась по данным наземных обследований на пробных площадях.

Анализ полученных картосхем, их верификация и уточнение по результатам наземного обследования позволили установить, что по космическим изображениям можно выделить четыре категории состояния древостоев: погибшие, сильно поврежденные, поврежденные и условно здоровые, при этом ошибка классификации составляет около 20%. Анализ временного ряда изображений установил существенное ускорение деградации лесов в конце 70-х–начале 80-х годов (рис. 5.6).

Прогрессирующее усыхание древостоев наблюдалось в юго-восточном направлении, в сторону преобладающего переноса эмиссий в течение вегетационного периода. В настоящее время граница повреждения распространяется более чем на 200 км от источников эмиссий, причем на расстоянии до 80–100 км древостой погиб. Распространение вредных выбросов в южном направлении ограничено горным массивом. При наземном обследовании пробных площадей южнее хребта повреждений не обнаружено; концентрация производных серы в хвое лиственниц также находилась в пределах фоновых значений.

Интенсивная гибель насаждений в конце 70-х–начале 80-х годов совпала с периодом пуска Надеждинского завода, когда интенсивность эмиссий SO_2 достигла своего максимума — 2,3 млн т/год, особую опасность представляли имевшие место залповые выбросы, сопровождавшиеся острыми повреждениями древостоев (пример обнаружения спутником EP-TOMS залпового выброса SO_2 Норильского комбината можно видеть на рис. 1.16, выброс показан стрелкой). Кроме того, выброс эмиссий через сверхвысокие трубы Надеждинского завода сделал возможным дальний перенос поллютантов через хребет Лонтокойский камень (высота труб завода, расположенных на возвышенности $h = 200$ м, составляет 250 м; трубы Никелевого и Медного заводов имеют меньшую высоту — 138–180 м и размещены на отметках порядка $h = 100$ м и ниже). На возможность дальнего переноса выбросов указывают также результаты дешифрирования материалов съемки, выполненной с помощью сканера МСУ-Э: шлейф выбросов Надеждинского завода просматривается на расстояние до 40 км, что приблизительно соответствует расположению максимальных отметок хребта Лонтокойский камень. Эти выводы подтверждаются данными дешифрирования

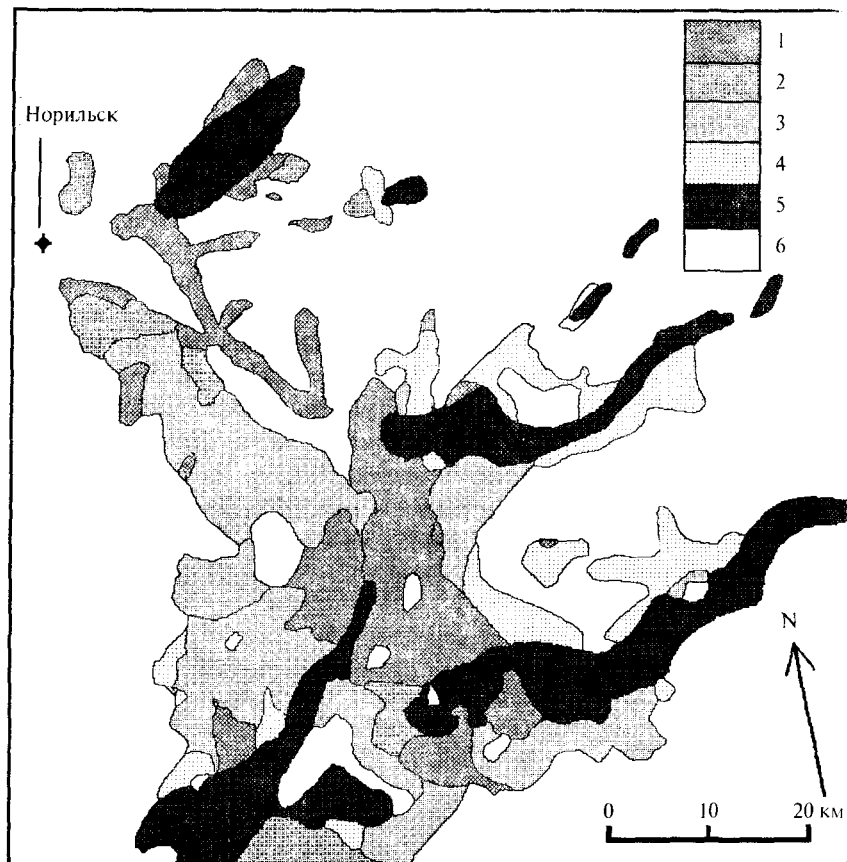


Рис. 5.6. Картосхема состояния притундровых лесов на 1984 г.

изображений, полученных сканером AVHRR. Съемка со спутников Landsat (MSS) использовалась для уточнения результатов дешифрирования мелкомасштабных снимков, а также для выбора наземных пробных площадей.

Проведенные исследования позволяют рассматривать воздействие двуокиси серы и ее производных как основную причину усыхания притундровых лесов. Следует отметить, что повреждающее воздействие оказывается именно через атмосферу, а не почву. Окислы азота, входящие в состав эмиссий, ввиду их сравнительно небольшого количества (~20 тыс. т/год), не могут существенно влиять на состояние древостоя; воздействие тяжелых металлов ограничено радиусом 30–40 км в окрестностях источников эмиссий.

Техногенный компонент деградации лесов обусловлен перестойностью насаждений, неудовлетворительным состоянием лесовозобновления, отчасти грибковыми заболеваниями. Кроме того, дестабилизирующее влияние на состояние притундровых лесов оказывает изменение климата. В сумме эффект факторов, не связанных с загрязнением, оценивается в 20%.

«Норильский феномен» представляет наиболее крупную техногенную катастрофу в зоне северных (бореальных) лесов.

5.4. КОНТРОЛЬ СОСТОЯНИЯ ВОДОЕМОВ

Космические средства контроля окружающей среды очень эффективны при изучении Мирового океана и внутренних водоемов. Для исследования Мирового океана созданы специальные спутники — «Океан-О» (Россия), MOS (Япония), SeaWiFS (США) и др. В частности, на спутнике NIMBUS-7 (США) был установлен сканер CZCS со спектральными каналами:

- 1-й — 0,433–0,453 мкм (голубой) для измерения хлорофилла в воде;
- 2-й — 0,51–0,53 мкм (зеленый) для измерения хлорофилла в воде;
- 3-й — 0,54–0,56 мкм (желтый) для измерения содержания желтого вещества в воде, для определения солености воды;
- 4-й — 0,66–0,68 мкм (красный) для оценки содержания аэрозоля в атмосфере;
- 5-й — 0,7–0,8 мкм (ближний ИК) для выделения суши и облаков;
- 6-й — 10,5–12,5 мкм (дальний ИК) для определения температуры.

Океан — это не только кладовая ресурсов, но и один из главных элементов общей системы экологического равновесия на Земле. Он сильно влияет на состояние атмосферы, формирование погоды, климат, энергетический и газовый балансы планеты, круговорот веществ в природе.

В Институте биофизики СО РАН и Красноярском государственном техническом университете выполняется цикл работ по программе «Зеленая волна» [10], в соответствии с которой разработаны методы мониторинга пространственной структуры, состояния и продуктивности морских и пресных водоемов, а также биоценозов суши на основе измерения распределения фитопигментов и биомассы в биосфере. Предложены новые эффективные методы измерения содержания хлорофилла и других фитопигментов на планете, позволяющие оперативно определять продуктивные зоны, следить за их динамикой. Используются информация со сканеров CZCS и AVHRR, при дешифрировании космических данных учитывается подспутниковая информация с тестовых участков, применяется атмосферная коррекция.

На первом этапе обработки спутниковых данных пиксели со сходными спектрами объединяются в классы по критерию минимальных расстояний (4.18), при этом выделяются основные объекты суши, воды и атмосферы. Далее оцениваются характеристики природных вод: программные средства используют алгоритмы многомерного статистического анализа и учитывают оптические свойства вод. В частности концентрация хлорофилла в воде определяется по уравнениям регрессии, связывающим концентрацию и яркость водной поверхности в разных спектральных каналах. Подобным же образом формируется информация о концентрации минеральных взвешенных веществ в воде, термодинамической структуре вод, интенсивности волнения и других характеристиках, описывающих состояние водных экологических систем. Найденные характеристики лежат в основе классификации типов вод непараметрическими статистическими методами классификации с обучением.

Разработанная информационная система мониторинга позволяет по первичной продуктивности оценивать динамические процессы в водных экосистемах и выдавать диагноз отклонений от среднемноголетнего для данного района и времени года; своевременно регистрировать и отслеживать динамику и развитие последствий экологических катастроф; оценивать антропогенное воздействие на живую оболочку Земли; прогнозировать изменения экологических систем на основе обучения модели по циклической динамике.

В качестве примера рассмотрим результаты спутникового мониторинга пространственно-временной изменчивости температуры поверхности Тихого океана. Температура воды в океане — наиболее важная характеристика поверхностных водных масс. Распределение температуры воды определяется не только зональными климатическими особенностями, но и формированием высокоградиентных температурных зон (температурных фронтов) [15] из-за взаимодействия различных структур течений, круговоротов и вихрей.

Данные о состоянии крупномасштабных полей температуры поверхности океана (ТПО) и температурных фронтов позволяют обнаружить зоны аномального нагрева или выхолаживания вод, что важно при оценке энергообмена между океаном и атмосферой. Кроме того карты фронтальных зон и данные о горизонтальных градиентах ТПО необходимы для оценки действия различных гидродинамических процессов в океане, таких как ветровые нагоны, синоптические вихри, меандрирование течений.

Важным свойством фронтальных зон является формирование гидробиологических неоднородностей в океане. Фронтальные зоны могут влиять на распределение планктона и рыб или служить своеобразным

жидкой границей, существующей длительное время и ограничивающей взаимодействие с другими типами вод и соответственно с биологическими структурами.

В северной части Тихого океана выявлена фронтальная зона, устойчивая уже в течение шести лет и образованная течениями Ойясио и Куроисио (рис. 5.7). Размеры этой зоны около 2000 км в поперечнике и 4000 км в длину. Средние температурные градиенты составляют 0,7–0,8 К/км. Для сезонной изменчивости этой фронтальной зоны характерна структура в виде прерывистых фронтальных зон размерами порядка 1000 км. Сезонная изменчивость в этом районе проявляется в размывании границ фронтальной зоны весной и осенью из-за уменьшения разности температур водных масс Ойясио и Куроисио и ослабления процесса переноса вод. При переходе от холодного сезона к тепловому в Беринговом море формируется фронтальная зона, ось которой может иметь как зональное, так и меридиональное направление.

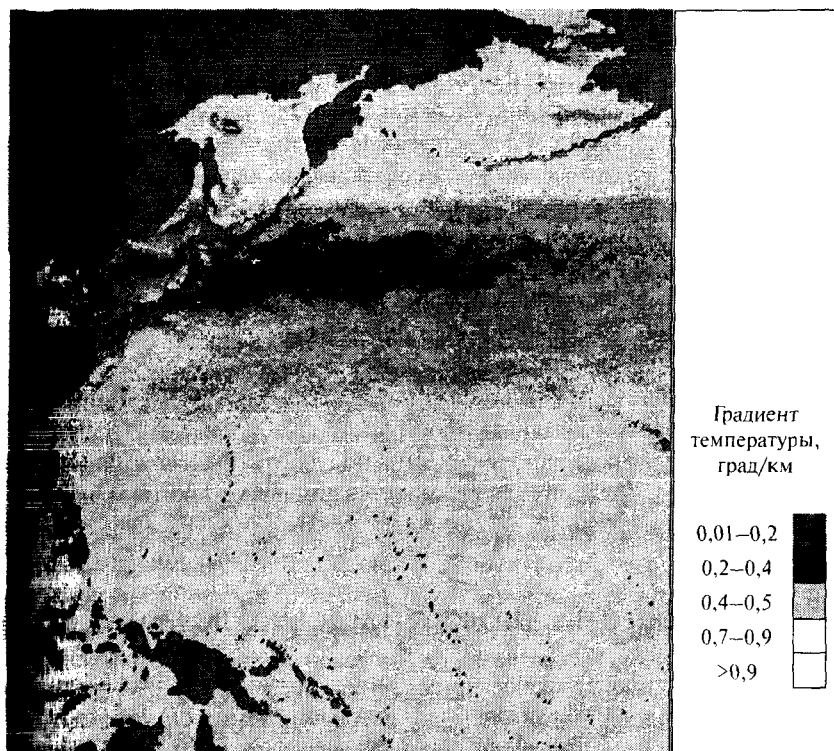


Рис. 5.7. Поверхностная температурная фронтальная зона в Тихом океане, найденная по решенным за 6 лет градиентам ТПО

По спутниковым данным проведены значительные по объему исследования сезонной и межгодовой динамики концентрации хлорофилла в поверхностном слое Мирового океана в глобальном и региональном масштабах. Это, в частности, позволяет находить зоны, перспективные для океанического рыболовства.

Разработанные программные средства используются для выявления динамики гидробиологических и гидрологических характеристик Красноярского водохранилища и Енисея с помощью спутниковых измерений. Эти средства можно также применять для исследования динамики пигментов фотосинтеза наземной растительности и решения практических задач при диагностике антропогенного воздействия крупных энергетических комплексов и промышленных предприятий на окружающую среду, при мониторинге других природных динамических процессов, таких как лесные пожары, весеннее снеготаяние и паводки.

На основе разработок Института биофизики СО РАН в Красноярском государственном техническом университете созданы пакеты прикладных программ «Зеленая волна» и «NOA» для обучения студентов основам экологического мониторинга водных объектов. Программа «Зеленая волна» использует карты распределения концентрации хлорофилла на поверхности океана, полученные путем усреднения по квадратам размерностью 20×20 км данных за различные периоды работы сканера CZCS. В процессе анализа необходимо выявить пространственно-временную структуру распределения концентрации хлорофилла на выделенной территории, найти участки с относительно устойчивым типом динамики концентрации.

Программа «NOA» предназначена для анализа температурных полей океанической поверхности и использует карты распределения температуры, полученные путем усреднения по квадратам размерностью 18×18 км спутниковых данных со сканеров CZCS и AVHRR. В процессе анализа необходимо выявить пространственно-временную изменчивость температуры поверхности океана, построить последовательность изображений, на которой прослеживается динамика фронтальных зон и фронтальных разделов.

5.5. КОНТРОЛЬ СНЕГОВОГО И ЛЕДОВОГО ПОКРОВА

Важнейшее практическое применение спутниковая информация находит в гидрологии и гляциологии при наблюдении снегового и ледового покрова, при мониторинге паводковых явлений, при обеспечении судоходства в арктических морях. Особенно это актуально для Си

Сибирь в районах с малой плотностью населения и редкой сетью метеорологических и гидрологических постов.

Обнаружение техногенного загрязнения местности. Снеговой покров зимой и в начале периода снеготаяния — удобный индикатор техногенного загрязнения местности. Например, при исследовании деградации лесов в Норильском промышленном районе выявлено соответствие между ореолом загрязнения снегового покрова, обнаруженного по снимкам космическим, и концентрацией тяжелых металлов (Cu, Ni, Co) в почве, которая в верхних почвенных горизонтах (0–10 см) превосходит фоновый уровень в 10–1000 раз в радиусе до 40 км от источника эмиссий [14].

Определение площади заснеженности. Важной задачей наблюдения снегового покрова является оперативное определение площади заснеженности, что необходимо для прогноза стоков рек со снеговым питанием, типичным для севера Сибири и для горных рек. В качестве примера рассмотрим результаты исследований, проведенных с использованием информации со спутника NOAA по бассейну Нижней Тунгуски в 1995 г. [16, 17]. Эта река со снеговым питанием отличается экстремальными перепадами уровня воды, которые могут достигать 20 м и более.

При обработке весенних изображений бассейна Нижней Тунгуски прежде всего было необходимо разделить пиксели на три вида: а) содержащие изображение снега (льда); б) относящиеся к изображению облаков; в) содержащие изображения почвы и растительности. Снег и облака обуславливают наиболее яркие фрагменты изображения в видимом участке спектра (1-й канал AVHRR — 0,58–0,68 мкм), причем облака имеют приблизительно одинаковую яркость в видимом участке и в ближнем инфракрасном диапазоне (2-й канал — 0,725–1,0 мкм). Яркость снега (льда) во 2-м канале меньше, чем облаков, и может составлять менее 60% их яркости. Изображения почвы и растительности имеют во 2-м канале большую яркость, чем в 1-м.

Для разделения пикселей на три указанных вида целесообразно использовать вегетационный индекс $NDVI$ (1.12), что позволяет в удобной форме выделять снег (лед) на фоне облаков и участков, свободных от снега (льда). Значениям $NDVI$ ставились в соответствие условные цвета: участки, содержащие снег и лед, на экране окрашивались в голубой цвет, облака — в белый, участки с изображениями почвы и растительности — в красновато-бурый. Это дает возможность уверенно различать снег, облака и свободную от снега поверхность. При этом талый, насыщенный водой снег выглядит серо-синим или даже сине-черным. Четко выделяется рисунок рельефа: в долинах, где теплее, мокрый снег более темный, чем сухой снег на водоразделах, который выглядит здесь светлее. В долинах снег сходит раньше, поэтому красно-бурые участ-

ки сначала появляются на дне долин. В ходе таяния они постепенно разрастаются, поднимаясь на возвышенные водораздельные пространства. Облака верхнего яруса (перистые и перисто-слоистые), отличающиеся незначительной водностью и почти прозрачные для солнечной радиации, как бы накрывают участки изображения поверхности земли беломатерчатой прозрачной вуалью, через которую «проглядывают» голубой снег и красно-бурые талые участки. Сплошные слоисто-кучевые облака с высокой водностью и значительной вертикальной мощностью изображаются в виде рисунка белого сравнительно однородного тона.

Для нахождения площади заснеженности использовалась следующая методика. Оператор выбирал фрагменты изображения, на которых интересующая площадь свободна от облаков. Для создания обучающих выборок задавались два типа эталонных участков: заведомо заснеженный и заведомо свободный от снега (например, расположенный ниже по высоте или южнее исследуемого района). Для каждого типа эталонов и каждого спектрального канала подсчитывались среднее значение яркости пиксела и стандартное отклонение яркости. Таким образом были получены восемь характеристик эталонов: средние значения яркости для снега (льда) и для фона (почва, растительность) в 1-м и во 2-м каналах, а также стандартные отклонения в 1-м и во 2-м каналах.

Для автоматического разделения изображения в пределах выделенного водосбора на участки двух классов — заснеженные и свободные от снега — использовался алгоритм (4.17). Заключительная процедура состояла в подсчете числа пикселей, относящихся к заснеженным участкам в пределах водосбора, и в вычислении площади заснеженности.

Площадь рассматриваемого бассейна Нижней Тунгуски (от поселка Ербагачен до поселка Тура ниже впадения реки Тембенчи) составляет 209 500 км². Определить заснеженность одновременно для всей этой площади не всегда возможно из-за наличия облачности. Поэтому площадь разбивалась на части размером примерно 100 × 100 км и определялась заснеженность раздельно для каждой из частей, не закрытых облаками, предварительно для каждой части было установлено распределение площади по высотным зонам в форме таблиц и графиков (гипсографические кривые), а также составлено ее ландшафтное описание. По описанию выявлялись участки, близкие по характеру ландшафтов и, следовательно, по условиям снегонакопления, если, конечно, они близки и территориально.

Для оценки заснеженности участков бассейна, закрытых облаками используется метод аналогии: 1) выбирается открытый участок-аналог по которому по рассмотренной методике оценивается степень заснеженности; 2) для этого участка по гипсографической кривой определяется высотное положение снеговой линии; 3) по найденной таким обра-

этой высоте снеговой линии и гипсографической кривой закрытого об-
 ластками участка определяется искомая площадь заснеженности. После
 восстановления всех пропущенных из-за наличия облачности данных вы-
 числяется заснеженность всего рассматриваемого бассейна. Время, затра-
 чиваемое на определение заснеженности (с учетом получения спутнико-
 вой информации в реальном времени, обработки и анализа изображений
 и автоматизированного расчета заснеженности), не превышает 1 час.

Рассмотренная методика определения площади заснеженности с ис-
 пользованием информации со спутников NOAA применена в техноло-
 гической схеме оперативного прогноза ежедневных уровней воды.
 Для расчета уровней воды на основе усредненных за 5 лет данных была
 построена модель стока Нижней Тунгуски и ее притока Тембенчи. В ка-
 честве подстроечных параметров модели использовались данные о за-
 пасах воды в снежном покрове перед началом снеготаяния по действу-
 ющим пунктам снегосъемок; приближенная оценка на их основе средних
 значений снегозапасов в бассейне; среднесуточная температура возду-
 ха и ее суточные суммы по ежедневным данным единственной функци-
 нирующей гидрометеостанции в поселке Тура; прогностические зна-
 чения этих характеристик на 3–7 суток вперед; ежедневные (на 8 часов
 утра) уровни воды Нижней Тунгуски в пунктах Ербагочен и Тура.

Космическая информация использовалась как важнейший подст-
 роечный элемент модели. Ежедневно ожидаемые данные о заснежен-
 ности сравнивались с текущей космической информацией; если расхож-
 дение превышало 3–4%, то в интерактивном режиме осуществлялась

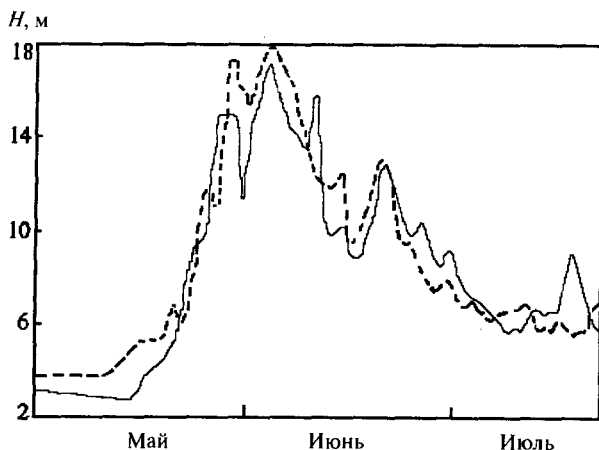


Рис. 5.8. Фактические (—) и прогнозируемые с заблаговре-
 менностью 4 суток (----) уровни воды р. Нижней Тунгуски за
 период весеннего половодья в 1995 г.

корректировка модели. Компьютер выдавал значения прогнозируемых уровней воды в поселке Тура за 2–7 суток. Такая методика полностью себя оправдала (рис. 5.8).

Обнаружение ледовых заторов на реках. Используя космическую информацию, можно своевременно обнаруживать опасные природные явления на реках. В Институте леса СО РАН и Красноярском государственном техническом университете накоплен большой опыт по обнаружению заторов на Лене, Енисее и их притоках (рис. 5.9). Как и при определении площади заснеженности, подсчитывался вегетационный индекс *NDVI* на интересующие районы. Это позволило различать облака, снег/лед, свободную ото льда поверхность воды и участки суши. На космических изображениях хорошо просматривались затопленные площади. Эти данные, немедленно передаваемые в органы МЧС, незаменимы при мониторинге чрезвычайных ситуаций в малонаселенных районах и служат основой для планирования разведывательных полетов авиации и мероприятий по уничтожению заторов. С использованием данных Красноярской станции HRPT проводились спасательные работы во время весенних наводнений в малонаселенных районах Якутии и в Красноярском крае.

Контроль ледового покрова. Получению космических изображений в оптическом диапазоне сильно мешает облачность. С этой точки зрения перспективно применение радиолокационных изображений с искусственных спутников Земли [18]. В условиях Сибири пассивные и особенно активные радиоволновые методы целесообразно использовать для оперативного контроля кромки ледостава на реках и водохранилищах в осенний, зимний и весенний периоды, для своевременного



Рис. 5.9. Ледяной затор в устье Ангары 3 мая 1995 г.

по обнаружения и картирования загорно-заторных явлений, для обнаружения разливов рек, пленки нефтепродуктов на поверхности водоемов.

Одной из важнейших задач, которую можно успешно решать методами активной радиолокации, является ледовая разведка в арктических морях. Судоходство по Северному морскому пути связано с плаванием судов в сложной ледовой обстановке. Навигационный период обычно проводится при неблагоприятных погодных условиях — частых снегопадах и туманах, густой облачности. Это затрудняет использование информации и применение спутниковых данных оптического диапазона при разработке безопасных маршрутов караванов судов.

Даже при снегопаде, тумане, облачности на радиолокационных изображениях хорошо просматриваются участки, свободные ото льда. На них среди льдов любого возраста отчетливо видны каналы и разводья. Различные по своим параметрам льды по-разному воспроизводятся на изображениях, имеют свои отличительные особенности и характерный рисунок. Ряд особенностей ледяного покрова выявляется только при радиолокационном обзоре и не обнаруживается в оптическом диапазоне. Радиоволны без существенного ослабления проникают сквозь снежный покров толщиной 40 см и более, если он не насыщен водой. Под слоем снега четко просматривается структура льда, выделяются замерзшие каналы и разводья, а также некоторые другие особенности ледяного покрова. Однако определение важнейшей характеристики ледяного покрова — его возраста — наиболее сложный элемент радиолокационной ледовой разведки, так как возраст можно оценить лишь по косвенным признакам.

5.6. ИССЛЕДОВАНИЕ ВЕРТИКАЛЬНЫХ ПРОФИЛЕЙ АТМОСФЕРЫ

К важным достижениям космического дистанционного зондирования относится оперативное определение поля вертикальных профилей земной атмосферы (параметры: температура, влажность и др.). Соответствующие исследования проведены во многих странах, в том числе в России [19, 20]. В настоящее время аппаратура TOVS для исследования поля вертикальных профилей атмосферы установлена на спутниках серии NOAA.

Радиометр TIROS Operational Vertical Sounder (TOVS) представляет собой комплекс из трех независимых систем, данные которых могут быть использованы для восстановления вертикальных профилей температуры, влажности воздуха, поля ветра, концентрации озона и других параметров:

20-канальный сканирующий ИК-спектрофотометр HIRS/2I имеет 19 спектральных каналов в инфракрасной области спектра и один в видимой, разрешение в подспутниковой точке 17,4 км, полоса обзора 2600 км;

3-канальный сканирующий стратосферный ИК-спектрофотометр Stratospheric Sounder Unit (SSU) имеет три спектральных канала в полосе поглощения углекислого газа, разрешение в подспутниковой точке 147,4 км, полоса обзора 1470 км;

4-канальный сканирующий радиометр Microwave Sounder Unit (MSU) микроволнового диапазона 50–58 ГГц имеет четыре канала в полосе поглощения кислорода, разрешение в подспутниковой точке 109,0 км, полоса обзора 2350 км.

В состав программного обеспечения Красноярской станции HRP1 входит пакет вертикального зондирования атмосферы TOVS ITPP версии 3.0. Пакет содержит программы выделения из потока данных информации TOVS, радиометрической калибровки, географической привязки, собственно определения вертикальных профилей атмосферы. С помощью этого пакета можно вычислить вертикальное распределение температуры воздуха, температуру точки росы, значение общего содержания озона в атмосфере, температуру и давление на поверхности Земли и на верхней границе облачности. Поле вертикальных профилей атмосферы определяется на неравномерной сетке точек (до 600). Объем получаемых данных с одного прохода спутника составляет около 250 кбайт.

Расположенные на различной высоте слои атмосферы имеют разную температуру и давление, при этом вертикальное распределение концентрации некоторых газовых составляющих, например CO_2 и O_3 , можно считать известным и неизменным. Для измерения характеристик атмосферы используют методы спектроскопии газов. Элемент атмосферы излучает, поглощает, рассеивает радиацию во всех направлениях. Интенсивность и длина волны линий в спектрах ряда газов, например CO_2 , зависит от параметров состояния элемента атмосферы (температуры, давления, концентрации газов). Производя измерения на различных длинах волн внутри конкретной полосы поглощения, удается выполнять вертикальное сканирование атмосферы в определенном диапазоне высот, т.е. измерять высотные профили температуры или удельной концентрации газовых составляющих атмосферы (в зависимости от выбранной полосы поглощения) [19, 20].

Совокупность процессов излучения, поглощения, рассеяния радиации элементом атмосферы описывается уравнением радиационного переноса, которое в общем виде имеет сложный характер. В ИК-диапазоне можно пренебречь рассеянием и упростить уравнение

Принимая ряд других упрощающих предположений (плоскопараллельная безоблачная атмосфера в состоянии термодинамического равновесия, независимость коэффициента теплового излучения ϵ подстилающей поверхности от длины волны, пренебрежение эффектом переизлучения атмосферой радиации от подстилающей поверхности), можно записать уравнение радиационного переноса в виде [19]

$$I(\lambda, \varphi) = \epsilon B[\lambda, T(p)] g[\lambda, T(p), q(p), \varphi] \Big|_{p=p_0}.$$

$$- \int_0^p B[\lambda, T(p)] T \{ \partial g[\lambda, T(p), q(p), \varphi] / \partial p \} dp,$$

где $I(\lambda, \varphi)$ — интенсивность излучения в интервале длин волн $(\lambda, \lambda + \Delta\lambda)$ на верхней границе атмосферы (давление $p = 0$), регистрируемая датчиком космического аппарата, $B(\lambda, T)$ — функция излучения Планка согласно (1.1), $T(p)$ и $q(p)$ — профили температуры и концентрации поглощающих компонентов в зависимости от давления p , $g[\lambda, T(p), q(p), \varphi] = g(\varphi, p)$ — функция пропускания вверх на уровне давления p для угла визирования φ с вертикалью в точке измерения, p_0 — давление на уровне подстилающей поверхности.

Первое слагаемое этого уравнения описывает излучение подстилающей поверхности (уровень p_0), ослабленное поглощением в атмосфере, второе — собственное излучение атмосферы, прошедшее через вышележащие слои и являющееся суперпозицией излучения разных слоев атмосферы.

Приведенное уравнение или его модификации лежат в основе интерпретации спектральных измерений, используемых для дистанционного зондирования атмосферы. Вид функций пропускания, входящих в уравнение переноса, определяется характеристиками конкретных спектральных линий или полос: интенсивностью, формой и шириной контуров линий.

Количественная информация о распределении основных метеорологических параметров получается в результате интерпретации спектральных измерений в полосах поглощения углекислого газа (линия 4,26 мкм, полоса 15 мкм), водяного пара (6,3 мкм, 18 мкм, 0,8 см, 1,35 см). Предположение о том, что вертикальное распределение CO_2 и O_2 известно и неизменно, позволяет разделить переменные в выражении для уходящего излучения, зафиксировать все параметры, кроме температуры, и найти профиль температуры.

Наличие облачности усложняет задачу. Для ее успешного решения используются новые длины волн, совершенствуется методика измерений

[19]. Если на спутниках первого поколения измерения проводились в носовом диапазоне поглощения CO_2 на 15 мкм и в окне прозрачности атмосферы на 8–12 мкм, то в настоящее время наряду с расширением измерений в ИК-диапазоне привлекается радиометрия в микроволновом диапазоне.

Главная трудность нахождения функций $T(p)$ и $q(p)$ заключается в том, что процедура обращения полученного интегрального уравнения относится к классу некорректных задач математической физики. Обратный оператор задачи неограничен, а решение является неединственным и неустойчивым по отношению к возмущениям. Поэтому для нахождения решений, адекватных реальному состоянию атмосферы, привлекается априорная информация, в качестве которой используются реперные среднеклиматические профили температуры $T_0(p)$ и влажности. В частности, в пакете TOVS ITPP версии 3.0 использована стандартная модель атмосферы NOAA, применяемая в США.

Уравнение относительно температуры нелинейное. Для его линеаризации можно искать не $T(p)$, а разность $T(p) - T_0(p)$. Полученные решением обратной задачи профили температуры и влажности (рис. 5.10) являются основой для вычисления других параметров атмосферы.

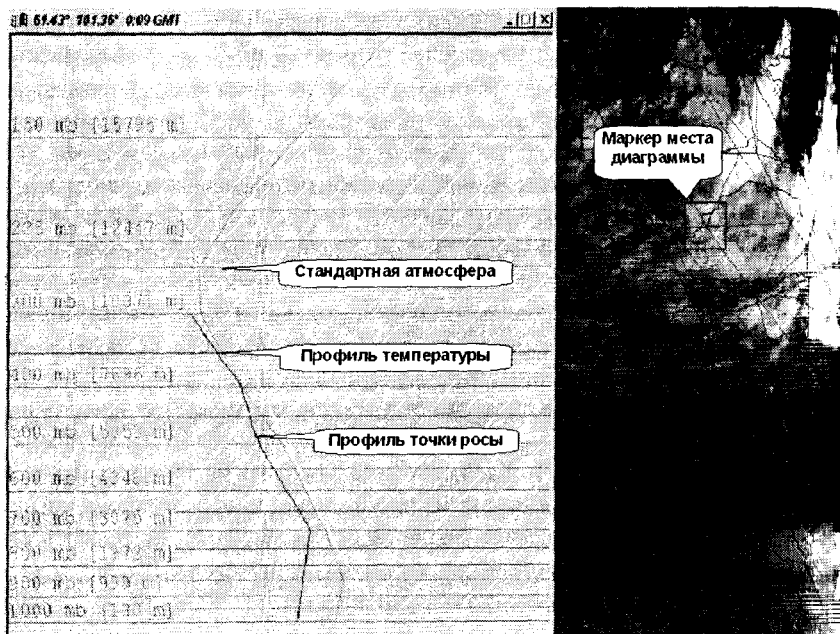


Рис. 5.10. Профиль температуры и влажности, полученный с помощью HIRS

Интегрируя уравнение состояния влажной атмосферы при известном профиле температуры, получают профиль геопотенциала — потенциальной энергии тяготения, отнесенной к единице массы. При этом считается, что ускорение силы тяжести различно в разных точках на поверхности и над поверхностью Земли [22]. Геопотенциальной поверхностью поля силы тяжести принято называть поверхность, перемещение по которой в любом направлении не требует затраты работы против этой силы. В метеорологии единицей геопотенциала считается геопотенциальный метр (гпм), который численно приблизительно равен обычному метру, но имеет размерность [Дж/кг]. При обработке данных HIRS в качестве характеристики влажности воздуха определяется зависимость точки росы и ее дефицита от высоты. В метеорологии точка росы — это температура, до которой должен охладиться водяной пар, чтобы достичь состояния насыщения при данном влагосодержании и неизменном давлении, дефицит точки росы в °С характеризует скорость испарения воды.

По полю геопотенциала рассчитываются направление и скорость геострофического ветра — идеализированного равномерного и прямолинейного движения воздуха при отсутствии силы трения по изобаре — линий равного давления. Действительный ветер в слоях атмосферы, лежащих выше 1 км, может быть близок к геострофическому [22].

Результаты спутниковых исследований вертикальных профилей атмосферы представляют значительный интерес для метеорологии и авиации. С точки зрения лесопожарного мониторинга важно дистанционное определение температуры и влажности в приземном слое и на различных высотах над поверхностью. Альтернативным методом получения вертикальных профилей атмосферы является аэрологическое зондирование с помощью радиозондов — миниатюрных метеостанций, поднимаемых в баллонах в воздух и снабженных радиопередатчиком. Данные о перемещении радиозонда характеризуют скорость и направление ветра. Аэрологическое зондирование на территории Красноярского края проводится раз в сутки в 6–8 пунктах, тогда как со спутников можно получать информацию до 4 раз в сутки в сотнях точек, практически на всей территории края. Возможность дополнения данных аэрологического зондирования спутниковыми данными достаточно актуальна.

Представляет интерес сравнение спутниковых данных и данных аэрологического зондирования. Решение такого рода задач — необходимая часть дистанционных исследований, которые нуждаются в постоянной проверке их достоверности, точности, надежности. После запуска спутника обычно проходит продолжительный период верификации передаваемой информации, необходимой также в течение всего периода эксплуатации аппарата.

В частности, сравнивались спутниковые и аэрологические данные в трех пунктах аэрологического зондирования Красноярского края за летние периоды 1995 и 1996 гг.: вертикальные распределения температуры воздуха, геопотенциальная высота и дефицит точки росы, скорость и направление геострофического ветра. Исходя из характера сравниваемых данных, необходимо иметь в виду следующее. Во-первых, можно говорить только о разнице в показаниях между спутниковыми и аэрологическими данными, измерений радиозондами, поскольку точность некоторых параметров оставляет желать лучшего. Во-вторых, следует отметить временной интервал между пролетом спутника и запуском радиозонда, составляющий в использовавшихся данных от 30 минут до 2 часов и приходящийся на период максимального нарастания в динамике суточной температуры (утро). В-третьих, спутник выдает результаты измерения в конкретный момент времени, усредняя их по мгновенному полю зрения прибора, составляющему круг диаметром 17,4 км в надири и эллипс $29,9 \times 58,5$ км на краю полосы обзора, а профиль атмосферы с помощью радиозонда получают в течение несколько часов по мере его подъема, причем радиозонд в данный момент выдает параметры только для точки, в которой находится. В-четвертых, зондирование со спутника производится в случайно расположенных точках, с которыми в подавляющем большинстве случаев место запуска радиозонда не совпадает. Радиозонд за время работы может быть унесен ветром за десятки километров от места запуска. Хотя использовались спутниковые данные в ближайших к станциям аэрологического зондирования точках, тем не менее иногда отклонение места превышало 50 км.

Результаты сравнения измерений показали, что средняя разность по температуре воздуха равна по абсолютному значению $1-2^\circ\text{C}$ и только в стратосфере увеличивается до 4°C на уровне 100 гПа, что соответствует высотам около 20 км. Коэффициенты корреляции между значениями температуры воздуха по спутниковым и аэрологическим данным на всех уровнях (кроме 100 гПа) превышают 0,9. В тропосфере в 90% случаев температура воздуха, измеренная со спутника, отличается от данных радиозонда не более чем на 3°C .

Сравнение данных о вертикальном распределении влажности носит приближенный характер. Из-за несовершенства датчика влажности значения температуры и точки росы (дефицита), получаемые с радиозонда, зачастую имеют большие погрешности. Поэтому выбранные случаи зондирования тщательно анализировались, сомнительные данные исправлялись по правилам, изложенным в соответствующих методических указаниях по метеорологическому прогнозу. При этом значения дефицитов точки росы на разных уровнях сопоставлялись

численном облачности разных ярусов по данным приземных наблюдений метеостанций.

В первом приближении можно утверждать, что средняя разность дефицитов не превышает 2°C на всех уровнях. Но разброс этих значений достаточно велик, только в 2/3 случаев разность составила не более 2°C . При анализе замечено, что значения измеренных дефицитов практически совпадают при небольших значениях дефицита влажности, когда погода пасмурная или выпадают осадки. При ясной, малооблачной погоде различия могут быть большими.

Средняя разность в определении высоты (геопотенциала) основных изобарических поверхностей по абсолютной величине составляет от 20 гПм на уровнях 850–300 гПа (1,4–10 км) и увеличивается до 100 гПм на уровне 100 гПа. Почти на всех уровнях в 3/4 случаев разность не превышает 50 гПм. При определении геопотенциала по спутниковым и аэрологическим данным коэффициент корреляции невелик только на уровне 850 гПа, с высотой он быстро возрастает и достигает порядка 0,9.

Анализ больших отклонений геопотенциала показал, что практически все они совпадают с моментами прохождения через район измерения хорошо выраженных в барическом поле и в поле геопотенциала ложбин и гребней (их осевых частей) или центральных частей небольших циклонов и антициклонов. Эти различия, видимо, можно объяснить тем, что радиозонд дает значение геопотенциала в точке, тогда как спутник — значение, усредненное по довольно значительной площади, что приводит к сглаживанию экстремальных значений. В областях пониженного давления высота изобарических поверхностей по данным спутника завышается, в областях повышенного давления — занижается по сравнению с данными радиозонда. Эта разность тем больше, чем больше градиент барического поля.

Геострофический ветер рассчитывается по полю геопотенциала и, как следствие, включает в себя ошибки в определении и температуры воздуха, и геопотенциала, что приводит к значительным расхождениям с показаниями аэрологического зондирования. Разница в направлении ветра в некоторых случаях может достигать 180° , особенно в малоградиентных полях, когда скорость ветра невелика. Разница в скорости ветра составляет 4–7 м/с и особенно заметна при малых скоростях. При большой скорости ветра различия в направлении и значении скорости относительно невелики, что делает спутниковые данные пригодными для прогноза штормовых ветров.

2 октября 1996 г. по территории Сибири прошел глубокий циклон, ярко выраженными фронтами, которые обусловили выпадение ливневых осадков и штормовые ветры, усиливающиеся до ураганных, что

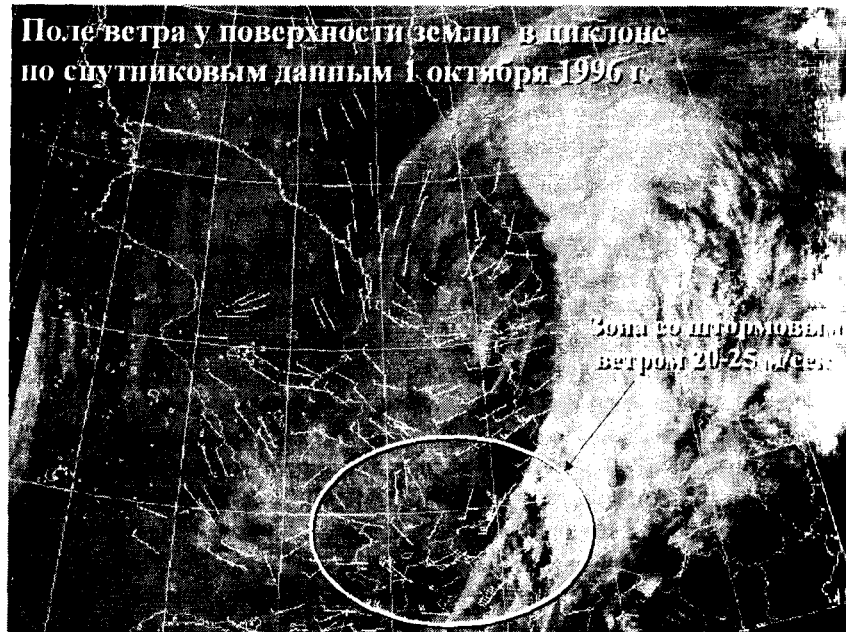


Рис. 5.11. Карта поля ветра

принесло значительный ущерб. На рис. 5.11 приведена карта поля ветра, полученная 1 октября, т.е. за сутки до циклона, наиболее пострадавший район отмечен овалом. Направление ветра, отображаемое флажками специальной формы, совпадает с направлением от штрихов флажка к его концу, а скорость ветра кодируется формой и числом штрихов: короткий штрих — 2,5 м/с, длинный — 5 м/с, треугольный — 25 м/с.

Спутниковые данные о состоянии атмосферы используются авиацией МЧС при планировании полетов в районы стихийных бедствий в Сибири.

5.7. ИЗУЧЕНИЕ ОБЛАЧНОСТИ

Спутники серии NOAA относятся к числу метеорологических и первоначально предназначались, в основном, для оценки глобальной и региональной метеорологической обстановки. Получаемые с их помощью изображения облачности позволяют идентифицировать и проследивать в глобальном масштабе различные синоптические объекты: циклоны, фронтальные системы, струйные течения, зоны конвергенции и др.

дискретных наземных наблюдений космические снимки выгодно отличаются тем, что дают целостную картину распределения облачности на достаточно больших территориях.

Изображения облачности можно использовать для оценки синоптической ситуации и уточнения распределения и эволюции целого ряда параметров состояния атмосферы, которые непосредственно со спутника не измеряются. Большую пользу приносят спутниковые наблюдения при анализе состояния погоды над районами с редкой сетью метеорологических станций. Редкая сеть ухудшает качество составляемых прогнозов погоды, зачастую не позволяя предвидеть опасные стихийные явления погоды. Космическая информация наиболее эффективна в сочетании с обычными (прямыми) наблюдениями.

За шесть лет эксплуатации Красноярской станции HRPT приема информации со спутников NOAA накоплен опыт в дешифрировании космических снимков для их использования в оперативной работе, разработаны методики применения космической информации для анализа сложившейся синоптической ситуации и уточнения прогнозов погоды, полученных от метеорологических организаций. Наибольший интерес представляли случаи или угрозы возникновения опасных или стихийных гидрометеорологических явлений на территории Восточно-Сибирского и соседних регионов.

К опасным гидрометеорологическим процессам и явлениям, которые могут спровоцировать природную чрезвычайную ситуацию (ЧС), относятся шквалы, смерчи, пыльные бури, метели, экстремальные снегопады и ливневые дожди, град. Они могут инициировать усугубление ЧС техногенного характера, а потому требуют особого внимания при анализе и прогнозировании погодных условий по космическим снимкам.

В любом случае кратко- или среднесрочный прогноз любого стихийного гидрометеорологического явления (СГЯ) на основе спутниковой информации сводится к комплексному анализу синоптической ситуации по данным сети наземных наблюдений Гидрометеослужбы и дистанционного зондирования Земли.

Общие принципы синоптического анализа снимков облачного покрова. В совокупности с синоптическими данными можно достаточно точно спрогнозировать возникновение, прекращение и интенсивность стихийных атмосферных явлений. Из опыта применения космической информации вытекают следующие принципы синоптического анализа снимков облачного покрова:

— облачность появляется не беспричинно, а при благоприятной для этого синоптической обстановке в определенных гидродинамических условиях, способствующих конденсации водяного пара. Поэтому при-

существование облаков над каким-либо участком земной поверхности не следует рассматривать как случайное малозначащее явление;

— тенденция развития синоптического процесса обнаруживается в поле облачности раньше, чем ее признаки появляются в поле температуры и давления.

С учетом этих принципов задачи синоптического анализа снимков облачного покрова сводятся к отысканию причин, обусловивших появление облаков в данном районе, и обнаружению признаков на изображении облачности, которые указывают на тенденцию в развитии синоптического процесса.

Анализ снимка облачного покрова целесообразнее всего начинать с оценки облачных образований крупного масштаба, а затем переходить к деталям облачного поля. Его можно условно разделить на следующие этапы:

— отыскание на снимке основных облачных систем, которые отражают движение воздуха крупного масштаба (облачные вихри, полосы);

— рассмотрение структуры и рисунка генеральных облачных образований (размеры, характер верхней поверхности облачности, резкости границ и т.д.);

— выделение облачности, которая образовалась внутри воздушной массы и непосредственно не связана с динамикой генерального синоптического процесса (кучевообразная облачность, туманы, слоистые облака и т.д.);

— выявление облачности, которая явно не связана с основными системами и не относится к внутримассовой, изучение ее структуры, строения верхней кромки, размеров и т.д. На эту облачность следует обращать особое внимание, поскольку она может быть результатом вновь зарождающегося процесса или признаком, указывающим на тенденцию в направлении перестройки синоптического процесса.

Выявление облачных вихрей и атмосферных фронтов. Для составления достоверного прогноза погодных условий большое значение имеет выявление облачных вихрей и атмосферных фронтов, с которыми в большинстве случаев связаны опасные и стихийные гидрометеорологические ситуации. Облачные вихри представляют собой крупномасштабные облачные образования с выраженной спиралевидной структурой. Фокус спирали — центр вихря. Облачные вихри отличаются большим разнообразием размеров и форм изображения в зависимости от стадии их развития, влажности воздуха, вертикальных движений, района формирования и др. На снимках (см., например, на рис. 5.12) наиболее отчетливо отображаются облачные вихри развитых и заполняющихся циклонов, имеющие, как правило, большое число спиралевидных облачных полос, сходящихся в одной точке.

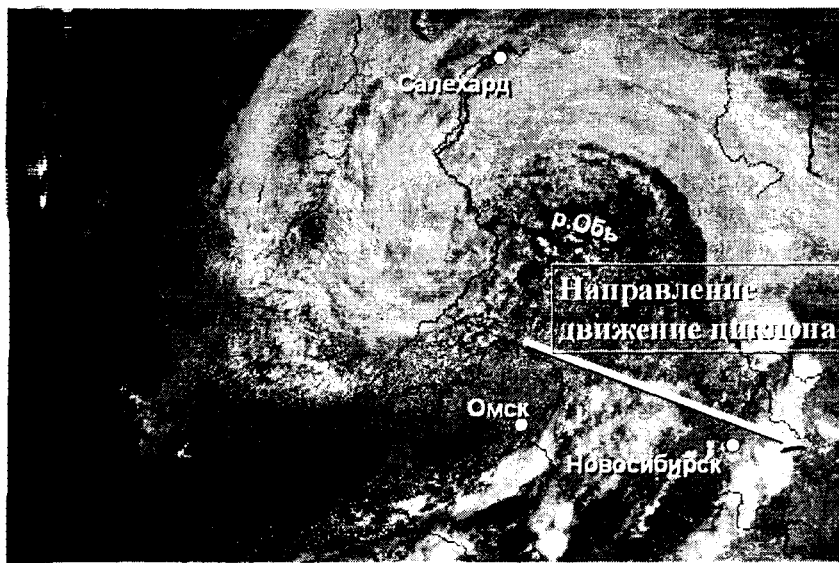


Рис. 5.12. Циклон над Западной Сибирью 5 июня 1998 г.

С облачными вихрями вплотную связаны облачные системы атмосферных фронтов, изображаемые на снимках в виде светлых полос различной ширины, яркости и структуры. Наиболее широкие и яркие полосы соответствуют активным фронтам с интенсивными восходящими движениями влажного воздуха, более узкие и менее светлые — неактивным, в области которых восходящие движения не получают должного развития. Во фронтальной полосе обычно присутствует не менее двух типов облачности. Сглаженные, ровные края указывают на преобладание слоистообразной облачности. Рваные, неровные границы являются признаком присутствия кучевообразной и кучево-дождевой облачности. Активность атмосферных фронтов уменьшается от центра циклонов к периферии, и это изменение активности обнаруживается на космических изображениях по уменьшению ширины полосы. Внешний вид облачности и конфигурация облачной полосы позволяют определить вид фронта на снимке. Это обстоятельство может служить основой для уточнения анализа синоптического положения в конкретном районе.

Облачные полосы холодных фронтов имеют четкую структуру в виде яркой полосы шириной 200—300 км и длиной более 1000 км, очень часто с вкраплениями округлых ярких пятен с резко очерченными краями. Полосы формируются из слоисто-дождевой облачности и отдель-

ных скоплений кучево-дождевых облаков. Активным холодным фронтам на снимках соответствуют непрерывные, хорошо развитые облачные полосы, яркие на снимках и в видимом диапазоне, и в инфракрасном. Линия холодного фронта у поверхности Земли практически всегда находится в пределах облачной полосы. У таких зон заметна циклоническая кривизна (прогиб в сторону теплого воздуха). Для холодного фронта характерны такие опасные и стихийные гидрометеорологические явления, как ураганы, шквалы, смерчи, пыльные бури, метели, экстремальные снегопады и ливневые дожди, град.

Теплый фронт (ТФ), как правило, хорошо выражен в поле облачности лишь в начальных стадиях развития циклона, поэтому распознавание его на снимках сложнее, чем холодного. Для теплого фронта наиболее типична облачная зона шириной 300–500 км и длиной до нескольких сотен километров в виде полос, причем длинные полосы встречаются редко. Облачная система ТФ имеет антициклоническую кривизну, вытягивается в сторону холодного воздуха и изображается в виде полосы. Положение облачной полосы ТФ обычно хорошо согласуется с положением приземной ложбины. При этом линию фронта у поверхности Земли следует проводить вблизи внутреннего края облачной полосы. Для теплого фронта характерны такие опасные явления погоды, как экстремальные снегопады и продолжительные интенсивные дожди.

Облачная полоса ТФ постепенно размывается и в процессе окклюзии сливается с облаками холодного фронта. Поэтому на снимках вместо ранее существовавшей облачной полосы ТФ заметен лишь незначительный выступ у точки окклюзии. Облачная зона, соответствующая фронту окклюзии (ФО), представляет собой плотную (яркую) полосу шириной около 300 км, обычно в форме спирали, внешне напоминающей гигантскую запятую, вершина которой находится в центре циклонической циркуляции на уровне облачности.

Облачные полосы, связанные с предфронтальной линией шквала, располагаются в теплом воздухе параллельно облачной полосе холодного фронта и часто бывают отдалены от нее относительно безоблачной зоной шириной 50–100 км. Наиболее интенсивные линии шквала наблюдаются там, где на снимках в видимом и инфракрасном спектрах видны яркие пятна кучево-дождевых облаков с наковальнями.

Облака зафронтальной линии шквала на снимках имеют вид прерывистых полос, расположенных в тыловой части облачных вихрей и состоящих из кучево-дождевых и кучевообразных облаков. В зависимости от интенсивности конвективных процессов ширина облачных полос составляет 50–200 км.

Кучево-дождевые облака. Основными признаками кучево-дождевых облаков являются: наиболее яркий (ярко-белый) тон изображения; отчетливо очерченные контуры облачности, хорошо различимые на фоне подстилающей поверхности и легко опознаваемые на фоне любой другой облачности; значительные колебания горизонтальных размеров; характерные выбросы (шлейф) наковален перистых облаков. Мощные кучево-дождевые облака, характеризующиеся высокой отражательной способностью, на снимках видимого и ИК-участков спектра имеют равномерную яркость.

Полезным для распознавания признаком кучево-дождевых облаков являются отбрасываемые ими тени, особенно в случаях, когда они проецируются сквозь толщу слоистообразных облаков. Кучево-дождевые облака без наковален выглядят яркими пятнами. Наличие их — хороший индикатор для прогноза гроз, ливней и шквалистых ветров в районе, для которого получена спутниковая информация.

Технология анализа и прогноза стихийных гидрометеорологических явлений. Она существенно зависит от имеющегося диагностического и прогностического материалов и способов их представления. Для составления качественного и достоверного краткосрочного прогноза синоптик должен получить из крупного метеорологического центра результаты крупномасштабных анализов и прогнозов для освещаемой территории, а также более детальных анализов, по крайней мере у поверхности Земли, для конкретных районов. При наличии данной информации космические снимки облачного покрова в видимом и инфракрасном участках спектра могут существенно облегчить задачу анализа и прогноза стихийных гидрометеорологических явлений.

При этом на первом этапе следует рассматривать крупномасштабные анализы и прогнозы, а также эволюцию крупномасштабных облачных систем за последние 6—12 часов и идентифицировать возможную «проблему дня»: приближение подвижной фронтальной зоны, попадание в зону стационарного фронтального раздела и т.д., т.е. определять возможность появления синоптической ситуации, в которой следует ожидать развития интенсивной конвекции, поскольку с ней связано подавляющее большинство случаев возникновения СГЯ.

Если такая «подозрительная» ситуация возникает, целесообразно провести более детальный анализ последней фактической информации: спутниковых, радиолокационных и обычных данных. При этом особое внимание необходимо уделить связи скоплений облачности с мезомасштабными синоптическими объектами: циклонами, фронтами, ложбинами и т.п. Кроме того, непрерывные в пространстве данные об облачности позволяют при анализе дискретных синоптических наблю-

дений заметить некоторые детали в структуре метеорологических полей, которые ускользают из поля зрения прогнозиста без этих данных.

Выделение мезомасштабных зон конвергенции, депрессий и связанных с ними очагов облачности завершает второй этап анализа, на котором определяются «подозрительные» с точки зрения конвекции облакты.

На следующем этапе производятся анализ и прогноз эволюции и смещения выделенных зон облачности, для которого целесообразно использовать следующие эмпирические правила:

- по фактическим или прогностическим высотным картам (либо по расчетным данным вертикального зондирования атмосферы TOVS, полученным со спутников NOAA) необходимо рассчитать среднюю скорость и направление воздушного потока и использовать их для прогнозирования будущего положения облачного массива (ОМ), поскольку после образования ОМ имеет тенденцию двигаться со средним потоком в слое с давлением 500–700 гПа (гектопаскалей);

- для оценки эволюции ОМ целесообразно использовать данные об изменении температуры верхней границы облаков, применяя для этого систему TOVS. При интерпретации ИК-изображений следует помнить, что осадки у Земли начинаются при температуре верхней границы облаков (ВГО) примерно -32° . При тенденции понижения температуры ВГО следует ожидать в ближайшие 2–4 часа увеличения интенсивности осадков и их площади, и наоборот;

- обострения зоны конвекции и усиления интенсивности развития в ней кучево-дождевых облаков следует ожидать при ее смещении в район более влажного и неустойчиво стратифицированного воздуха. Поэтому по имеющимся данным о температуре и влажности дистанционного вертикального зондирования TOVS следует оценить степень неустойчивости и влагосодержания атмосферы в районе, куда смещается ОМ.

Оценив в процессе предварительного анализа общие тенденции развития синоптических процессов в интересующем районе или регионе, направление и скорость смещения опасных зон и их эволюцию, можно переходить к анализу этих зон на предмет образования в них опасных и стихийных гидрометеорологических явлений.

Прогнозирование сильных шквалов и смерчей. Шквалы и смерчи относятся к локальным явлениям погоды, возникают внезапно (чаще после полудня), кратковременны (часто несколько минут) и охватывают сравнительно небольшие площади. Это определяет трудности их исследования и прогноза. Шквал — резкое усиление ветра до 25–30 м/с с изменением направления в течение короткого времени. Он всегда связан с развитием активной конвекции и сопровождается интенсив-

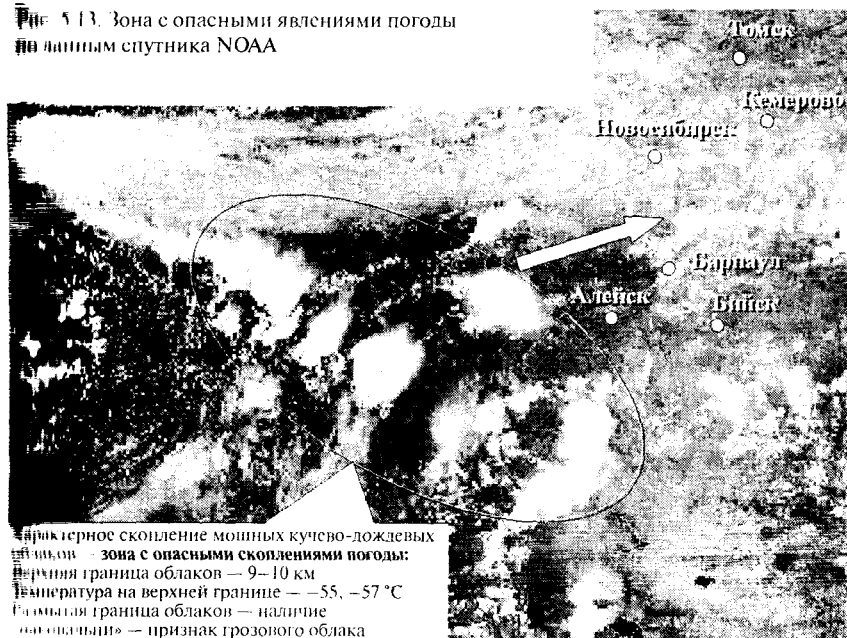
Шквалы осадками и градом. Внутримассовые шквалы в жаркую летнюю погоду, обусловленные мощными кучево-дождевыми облаками, не представляют серьезной опасности, скорость ветра не превышает 20 м/с. Фронтальные шквалы, образуемые на атмосферных, преимущественно холодных фронтах, представляют собой довольно опасное явление. Шквалы сопровождаются кучево-дождевыми облаками (Cumulonimbus, Сb) с низким основанием и высокой вершиной. Чем мощнее Сb, тем сильнее шквалы [21, 22].

Анализ облачности на космических снимках в дни с развитием шквалов разной интенсивности показал, что перед их возникновением в районах их развития или вблизи этих районов уже в ночные или предутренние часы выявляются характерные скопления Сb, имеющие следующие параметры:

- округлую, овальную, в редких случаях неправильную форму с четкими границами;
- линейные размеры скоплений в 90% случаев в пределах 100–150 км, в 10% случаев более 450 км и всегда более 100 км;
- если характерные скопления Сb появляются в дневное время, то с ними могут быть связаны только слабые шквалы.

На рис. 5.13 показано изображение облачности для случаев развития сильных шквалов. В данной ситуации скорость ветра составляла до 17 м/с.

Рис. 5.13. Зона с опасными явлениями погоды по данным спутника NOAA



Шквалы, особенно очень сильные, имеют много общего со смерчами. Смерчем называют вихрь с вертикальной осью вращения (в отличие от шквалов с горизонтальной осью), сопровождающийся колоссальными скоростями ветра. Большинство смерчей связано с линиями шквалов или активными холодными фронтами с грозами. Обычно при смерче из грозового облака отщепляется облачный столб в виде воронки. Скорость ветра внутри смерча может достигать 100 м/с и более. Условия образования смерчей наиболее благоприятны непосредственно на линии приземного фронта, вблизи поверхности Земли (это узкая полоса шириной около 50 км по обе стороны линии фронта).

К таким условиям относятся:

- наличие развитых мощных скоплений Сб, которые хорошо видны на снимках облачности с ИСЗ на территориях до 200—400 км. Верхняя граница Сб в большинстве случаев составляет более 13 км;
- наличие фронта с большими градиентами скорости ветра по высоте, способствующими образованию вихрей;
- наличие большой неустойчивости и большой доли влаги во всей толще воздушной массы.

За шестилетний период эксплуатации Красноярской станции НРПТ приема информации со спутников NOAA о возникновении смерчей на территории Восточной Сибири не было.

Прогнозирование сильных осадков в холодный период года. В умеренных широтах значимые осадки выпадают из слоисто-дождевых облаков с вкраплениями в них кучево-дождевых, в которых создаются условия для увеличения скорости падения водяных капель или ледяных кристаллов по сравнению со скоростью восходящих движений воздуха. Особо опасными считаются осадки при интенсивности более 20 мм за 12 часов. В холодный период года осадки такой интенсивности формируются на атмосферных фронтах при подъеме влажных теплых воздушных масс в циклонах. Продолжительные по времени и интенсивности восходящие движения способствуют накоплению больших запасов влаги в облачных системах, а механизм циркуляции воздушных частиц в облачных системах — к образованию и выпадению осадков. Очень сильные осадки (СГЯ) связаны с прохождением через пункт хорошо развитых циклонов преимущественно с юга или юго-запада [21, 22]. На космических снимках хорошо развитым южным циклонам соответствуют обширные облачные вихри диаметром более 500 км и с верхней границей облачности более 7 км. Сильные и очень сильные осадки никогда не отмечаются при перемещении воздушных масс с севера, северо-востока, востока и юго-востока.

В 100% случаев СГЯ на космическом снимке отмечены центральные части облачных вихрей. Более чем в половине случаев СГЯ отме-

чаются яркие широкие облачные полосы. Таким образом, опасные метеопадения можно связать с облачностью циклонов и атмосферных фронтов.

Прогнозирование осадков сильной интенсивности теплого полугодия.

К конвективным осадкам теплого полугодия относятся ливневые дожди и град. С сильными ливнями чаще всего связаны сели, паводки, подтопления, повреждения сельскохозяйственных культур и пр., а с градом – повреждения сельскохозяйственных и садовых культур, строения и автотранспорта.

Опасные и особо опасные конвективные осадки развиваются в основном в мезомасштабных облачных системах [21, 22]. Эти системы не всегда можно обнаружить в поле температуры и влажности. На космических снимках они появляются за 6–12 часов до осадков. Поэтому большое значение имеет методика прогноза опасных и особо опасных конвективных явлений с использованием всего комплекса синоптических, радиолокационных и спутниковых наблюдений.

Прогнозы всех указанных явлений можно составлять по следующей схеме:

- определение (расчет) траектории перемещения облачной системы на период прогноза;
- определение исходных параметров облачной системы вдоль траектории перемещения;
- определение эволюции параметров облачной системы;
- расчет прогнозируемого явления по существующим формулам или графикам, разработанным Росгидрометом.

Следует иметь в виду, что для образования града необходимо достижение верхней границы облака уровня естественной кристаллизации облага, в противном случае прогнозируется выпадение ливневого дождя. Вероятность выпадения града определяют высотой изотермы 0°.

Опасный и особо опасный град возникает в определенных облачных системах, выделяющихся на снимках с ИСЗ в скоплениях кучевых дождевых облаков с четкими границами овальной, круглой или грибовидной формы размерами более 300 км или в зонах конвективной облачности молодого циклона с активным холодным фронтом.

В течение 6 лет эксплуатации Красноярской станции HRPT приема информации со спутников NOAA накоплен опыт практического использования спутниковых наблюдений для уточнения прогнозов погодных условий на территории Восточно-Сибирского региона, особенно для обнаружения стихийных гидрометеорологических явлений, и прежде всего в труднодоступных районах, где традиционные методы не применимы.

5.8.1. Ультрафиолетовое излучение Солнца

Ультрафиолетовое (УФ) излучение составляет менее 5% общего солнечного излучения, однако его значимость в процессах жизнедеятельности очень велика из-за сильного воздействия на различные химические и биологические процессы. Интенсивность УФ-облученности зависит от высоты Солнца над горизонтом и максимальна в местных полдень.

С точки зрения биологического воздействия выделяют три области длин волн: *A* ($315 < \lambda < 400$ нм), *B* ($280 < \lambda < 315$ нм) и *C* ($\lambda < 280$ нм). Излучение области *B* обладает наибольшей эритемной эффективностью, т.е. наибольшей способностью вызывать эритему (расширение мельчайших капилляров кожи и сопутствующие этому процессы). Благоприятное в малых дозах, эритемное воздействие оказывается отрицательным в больших дозах, так как сильная эритема — это уже воспалительный процесс. Излучение области *B* обладает общим терапевтическим и антирахитичным действием.

Излучение в области *A*, прилегающей к видимой части спектра, менее эффективно биологически, но оказывает благоприятное эритемное и загарное воздействие. Излучение в области *C* обладает мощным бактерицидным действием и оказывает сильное негативное влияние на живые и растительные организмы. Интенсивность УФ-облученности зависит от высоты Солнца над горизонтом.

Основным глобальным фактором, ослабляющим УФ-излучение, является атмосферный озон, а также атмосферные аэрозоли — взвешенные в воздухе твердые частицы или капли жидкости природного и антропогенного происхождения. Частые лесные пожары в регионе Сибири сопровождаются выбросами дыма, который поглощает УФ-излучение. На интенсивность УФ-облученности у поверхности Земли сильно влияет облачность, в целом снижая прозрачность атмосферы. Но рассеянное УФ-излучение может иногда возрастать и при наличии облаков, что наблюдается, когда край кучевого облака «наплывает» на солнечный диск. В регионах с большим числом облачных дней, например в Западной и Восточной Сибири (до 80% облачных дней), влияние облаков в среднем даже более существенно, чем влияние озонового слоя, поэтому интенсивность УФ-излучения здесь не всегда напрямую зависит от количества озона в атмосфере.

Наиболее велико УФ-излучение в тропиках и субтропиках. На средних широтах на равнине, в частности в Сибири, наблюдается дефицит его, достигающий в среднем 30–45% медицински обоснованной нор-

По мере подъема в горы интенсивность увеличивается, так как излучение все меньше поглощается атмосферой. В тропиках слишком долгое пребывание на солнце без одежды и большая УФ-экспозиция могут привести к появлению раковой опухоли кожи — меланомы. Интересно, что чернокожие жители южных штатов США заболевают этой болезнью в 7 раз реже, чем белые, так как в их коже содержится пигмент, задерживающий УФ-излучение.

2.2. Экологическая роль озона

Озон, несмотря на малое его содержание в атмосфере, играет важную роль во многих биосферных процессах, поскольку является основным поглотителем солнечной ультрафиолетовой радиации в областях и т.д. Существенное уменьшение общего содержания озона (ОСО) соответственно увеличение УФ-излучения может привести к негативным последствиям и нанести ущерб народному хозяйству (снижение урожайности сельскохозяйственных культур, разрушение морских экосистем и др.) и здоровью населения (поражение зрения, рост числа заболеваний раком кожи и др.).

Озон распределен в атмосфере на высотах от 0 до 80 км. В тропосфере содержится менее 8% озона, на стратосферу приходится 80–85% его. Общая масса озона в атмосфере составляет около $3 \cdot 10^9$ т [22, 23].

Содержание озона в атмосфере относительно невелико и очень изменчиво. Геофизики выражают общее количество озона (ОСО) в столбе атмосферы — от 80 км до поверхности Земли — через толщину (в сантиметрах) того слоя, который образовал бы весь озон атмосферы, сжатый отдельно и приведенный к нормальному давлению 760 мм рт. ст. и нормальной температуре 0 °С. Значение 0,001 см называют единицей Добсона (сокращенно е.Д.). В среднем по земному шару ОСО составляет 290 е.Д., или 0,29 см.

С точки зрения экологии наиболее важен стратосферный озон. Задерживая основную часть УФ-излучения, он оказывается глобальным защитником всего живого на Земле. В тропосфере же озон играет скорее роль врага, чем союзника, являясь вредным для здоровья людей продуктом промышленного загрязнения воздуха. При взаимодействии с окружающей средой озон действует как мощный окислитель с сильным медико-биологическим воздействием. ПДК озона по стандартам России составляет 0,03 мг/м³ по среднесуточной концентрации и 0,16 мг/м³ по максимально разовой с первым классом опасности. Вклад приземного озона в общее содержание его в атмосфере невелик, но локальные концентрации могут достигать опасных для живой природы значений и в несколько раз превышать уровень ПДК.

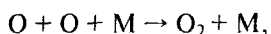
Солнечные лучи падают по нормали к экваториальной атмосфере, проходя в ней наименьший путь. На средних и высоких широтах солнечные лучи падают на атмосферу под углом δ к нормали и в соответствии с законом Бугера (см. п. 1.2.3 и рис. 1.8) проходят в ней больший путь. Обычно $\delta = 90^\circ - \psi$, где ψ — высота Солнца над горизонтом. В день весеннего и осеннего равноденствия 21 марта и 23 сентября в местный полдень $\delta = \varphi$, где φ — широта местности, в полдень в день летнего солнцестояния 22 июня угол $\delta = \varphi - \epsilon$ минимален. В день зимнего солнцестояния 22 декабря в полдень угол $\delta = \varphi + \epsilon$ максимален. Здесь $\epsilon = 23^\circ 26''$ — угол между осью вращения Земли и осью мира (т.е. осью, вокруг которой происходит кажущееся вращение небесной сферы).

Если X — общее содержание озона, то по закону Бугера эффективное общее содержание озона, ответственное за поглощение солнечного ультрафиолета, $X_2 = X / \cos \delta$. На широте 50° эффективное ОСО превышает X минимум в 1,12 раза, на широте $55-56^\circ$ (Москва, Новосибирск, Красноярск) X_2 превышает X не менее чем в 1,18 раза, на широте 69° (Норильск) не менее чем в 1,42 раза. Таким образом, если бы над Красноярском, Новосибирском или Москвой 22 июня наблюдалось аномально низкое для этого времени и этих широт ОСО в 270 е.Д., то в поглощении ультрафиолетового излучения фактически участвовал бы слой с ОСО $X_2 = 320$ е.Д., это вполне достаточно для подавления вредного влияния ультрафиолета В и С. 8 августа 2000 г. над Красноярском по данным спутника TOMS-EP ОСО действительно составляло 274 е.Д., но $X_2 > 380$ е.Д.

5.8.3. Образование и разрушение озона

Озон постоянно образуется в верхней стратосфере из молекулярного кислорода под действием УФ-радиации Солнца [22, 23]. Наибольшее количество озона возникает в стратосфере в результате фотохимических реакций близи экватора, где солнечные лучи пронизывают атмосферу под прямым углом.

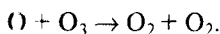
Процесс начинается с диссоциации молекулы кислорода на два атома под воздействием ультрафиолетового излучения Солнца и частиц высоких энергий, входящих в состав космических лучей. Образовавшиеся атомы кислорода либо соединяются вновь между собой в присутствии третьей молекулы М:



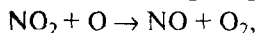
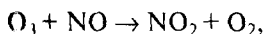
Атомы кислорода взаимодействуют с молекулой O_2 (также в присутствии третьего тела), образуя молекулу озона:



На озон, как и на молекулярный кислород, влияют внешние факторы (излучение, потоки частиц), под действием которых он разрушается (диссоциирует). Наконец, будучи химически активными, атомы O и молекулы O_3 могут взаимодействовать между собой (кислородный цикл):

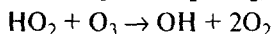
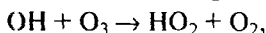
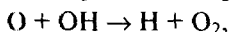
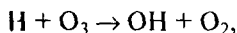


К разрушению озона также приводят реакции азотного цикла:

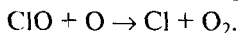
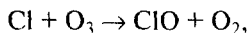


причем в результате достаточно высоких концентраций NO и NO_2 скорость гибели O и O_3 может стать во много раз больше, чем в предыдущей реакции. Эти реакции наиболее существенны в слое с максимумом содержания озона.

В нижней стратосфере на разрушение озона и атомарного кислорода сильнее всего влияют водород и его окислы (водородный цикл):



На озон заметно воздействуют присутствующие в стратосфере соединения хлора, который может участвовать более чем в 70 реакциях разрушения озона, причем окись хлора выступает в роли катализатора (хлорный цикл; подобно хлору ведет себя бром):



Источником атомов хлора в стратосфере является диссоциация его соединений под воздействием излучения Солнца.

Таким образом, к разрушению стратосферного озона приводят четыре фотохимических цикла: кислородный, азотный, водородный и хлорный. Правда, эффект их действия не равен сумме эффектов от отдельных циклов, так как химические реакции между членами различных семейств снижают воздействие циклов на озон. По одним оценкам, около 1,2% образующегося озона поступает в тропосферу, 17% разрушается при кислородном цикле, 11% — при водородном, остальные 70% — при азотном. По другим данным, 17% озона разрушается при кислородном цикле, 9% — при водородном, 70% — при азотном и 4% — при хлорном.

В принципе между процессами образования и разрушения озона должно быть динамическое равновесие: сколько озона возникает, столько его и разрушается. Тем не менее в 80-х — начале 90-х годов XX в. суммарное глобальное содержание озона ежегодно уменьшалось на 0,3–0,5%, что может свидетельствовать о нарушении равновесия. Однако в самом конце 90-х годов отмечен рост глобального содержания озона.

В 80–90-е годы XX в. многие ученые пришли к выводу, что хлор и его соединения могут быть причиной разрушения стратосферного озона. Развитие цивилизации приводит к росту выбросов хлорных соединений и их накоплению в атмосфере, одну из ведущих ролей в этом процессе играют так называемые фреоны. Доказательством этого, как утверждается, служит феномен «озоновой дыры» над Антарктидой.

Опасность дальнейшего роста выбросов хлорсодержащих соединений в атмосферу вызвала тревогу на международном уровне. По инициативе ученых ведущие государства подписали в 1987 г. Монреальский протокол, согласно которому должно быть резко сокращено использование наиболее опасных долгоживущих фреонов и других опасных для атмосферы веществ. К настоящему времени к протоколу присоединились около 170 стран. Образован специальный фонд помощи развивающимся странам для перехода на технологии, не использующие озоно-разрушающие вещества. 16 сентября объявлено Международным днем защиты озонового слоя. В Законе Российской Федерации «Об охране окружающей природной среды» есть статья 56, посвященная охране озонового слоя Земли. Мероприятия по защите озонового слоя — пример объединения усилий ученых, политиков и общественности в деле решения глобальных экологических проблем. Правда, нет пока оснований утверждать, что именно эта деятельность привела к глобальному увеличению содержания озона в конце 90-х годов, так как накопление фреонов в атмосфере продолжается.

Анализ влияния крупных промышленных комплексов Сибири на озоновый слой не обнаружил заметных изменений ОСО. Однако обширные лесные пожары приводят к некоторому уменьшению ОСО,

по спутниковым данным была обнаружена корреляция между интенсивностью лесных пожаров в Красноярском крае, республиках Хакасия и Тува и общим содержанием озона в регионе [24, 25]. Нападение стран НАТО на Югославию весной 1999 г. и авиационные налеты (до 500–700 самолетовылетов в сутки) вызывали уменьшение ОСО на Балканах. По данным со спутника TOMS-EP была выявлена корреляция между интенсивностью налетов и уменьшением ОСО, которое доходило до 10–12% [26]. Факт уменьшения ОСО был также отмечен наземными измерениями в Греции, территория которой непосредственно примыкает к зоне конфликта. Анализ влияния и лесных пожаров, и результатов авианалетов еще раз показывает, насколько ранним озоновый слой Земли.

5.8.4. Методы измерения общего содержания озона

Оптические свойства озона — поглощение и излучение им радиации — играют большую роль в физических процессах атмосферы. Они же используются для измерения ОСО. На поверхности Земли общее содержание озона в столбе атмосферы определяют, оценивая ослабление ультрафиолетового излучения Солнца в узких интервалах на двух (четырех) длинах волн с различным показателем ослабления озона. Это позволяет исключить влияние аэрозолей и других атмосферных помех, а также факторов, связанных с вариациями излучения на верхней границе атмосферы. В практике отечественной озонотрии получил распространение озонметр М-124 конструкции Г.П. Гушина, содержащий светофильтры для выделения двух участков спектра при $\lambda = 300$ и 326 нм. Оптическая схема озонметра обеспечивает следующие функции:

- наводку прибора на солнечный диск или на зенит неба с тем, чтобы излучение от диска Солнца или участка неба вблизи зенита попадало на фотоприемник под сравнительно небольшим телесным углом;
- выделение поочередно двух участков в ультрафиолетовой области спектра, из которых первый, наиболее коротковолновый, находится в полосе поглощения озона.

Определение ОСО производится путем сравнения показаний прибора в двух участках спектра по специальным таблицам и номограммам. Для М-124 основная погрешность, без учета систематической погрешности образцового средства измерения ОСО, составляет 3,2% при измерении по прямому солнечному свету и 5% при измерениях по облачному зениту неба.

За рубежом получили распространение озонотры Добсона и Брессера. На конец 80-х годов XX в. мировая сеть наземных озонотриче-

ских станций включала в себя 131 станцию наблюдений, которые очень неравномерно распределены по территории суши. Вероятность обнаружения с помощью этой сети аномалий в распределении озона мала, даже если аномалии имеют линейный размер в 1000 км.

Для глобального мониторинга озоносферы необходимы спутниковые данные об озоне, которые получают с помощью оптической аппаратуры, установленной на спутниках и работающей в УФ- и ИК-диапазонах по направлению в надир. В УФ-диапазоне методика измерения ОСО аналогична наземной: спутниковые озонметры регистрируют солнечное излучение, рассеянное назад молекулами озона. В спектрофотометре TOMS (США) для этого используются две пары длин волн: 0,3125 и 0,3312 мкм, 0,3175 и 0,3398 мкм, выделяемые монохроматором на дифракционной решетке. TOMS позволяет также определять концентрацию метана и двуокиси серы в атмосфере. Этот прибор успешно работал на российском спутнике «Метеор-3» и японском ADEOS, в настоящее время функционирует специальный озонметрический спутник TOMS/EP.

Оперативные данные об ОСО, полученные с помощью прибора TOMS, появляются в сети Internet по адресу <http://jwocky.toms.gsfc.nasa.gov> на следующий день после измерений, каждый файл содержит информацию об ОСО по всему земному шару с разрешением 100–120 км. Через месяц по адресу <http://daac.gsfc.nasa.gov/data/dataset/TOMS> приводятся откорректированные данные об ОСО с учетом отражений от подстилающей поверхности и атмосферы, поглощения атмосферным аэрозолем, взаимного расположения спутника, Солнца и пункта на поверхности Земли. Кроме того, на сайте NASA содержатся информация об ОСО над различными населенными пунктами и обширные архивные материалы. Погрешность определения ОСО прибором TOMS около 1–2%.

Прибор HIRS спутника NOAA регистрирует восходящее излучение не только в линиях поглощения углекислого газа и водяного пара, но и в полосе поглощения озона на $\lambda = 9,59$ мкм (рис. 5.14). При использовании пакета TOVS, кроме вертикальных профилей температуры и влажности, оценивается общее содержание озона в полосе 2000 км в 400–600 случайным образом расположенных точках, пространственное разрешение составляет 20–80 км. Случайный характер расположения точек представляет значительные неудобства в последующей интерпретации, поэтому целесообразно произвести пересчитывать данные с нерегулярной сетки на регулярную с использованием процедуры Криге. Окончательно данные об ОСО выдаются в виде карт изолиний или поверхностей, привязанных к географическим координатам. Однако на этой длине волны присутствует небольшое поглощение излучения во-

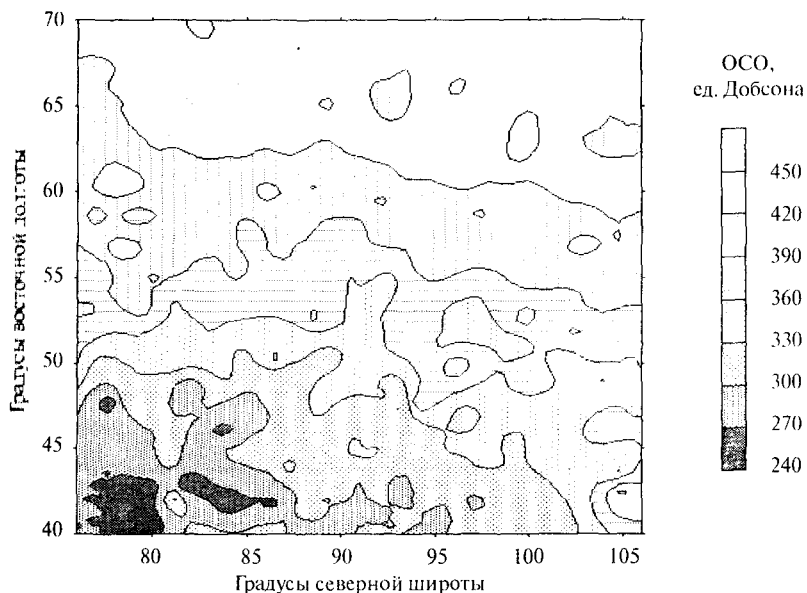


Рис. 5.14. Карта озонового слоя 25 апреля 1998 г. в 11 ч 25 мин. по Гринвичу: среднее $\langle \text{OCO} \rangle = 361$, стандартное отклонение — 55 единиц Добсона

длиным паром, и потому HIRS несколько завышает ОСО (в среднем, до 6%) [24].

5.8.5. Пространственно-временная динамика стратосферного озона

Кроме химических и фотохимических процессов, связанных с озоном, значительный интерес представляют пространственные и временные вариации содержания озона в атмосфере. К настоящему времени детально исследовано широтное распределение ОСО (табл. 5.1). Половина всего стратосферного озона сосредоточена в поясе от 35° с.ш. до 35° ю.ш. [22, 23].

ОСО испытывают суточные, сезонные, годовые и многолетние изменения. Среднее глобальное ОСО составляет около 290 ед.Д., наблюдаемая в природе толщина слоя озона меняется в широких пределах — от 90 до 760 ед.Д., а отклонение ОСО в течение суток от своего среднего значения может достигать до 25%. Наименьшие годовые изменения ОСО наблюдаются в экваториальной зоне, из которой воздушные течения весной переносят озон в средние и высокие широты. В этих ши-

**Таблица 5.1. Доля, %, атмосферного озона, содержащегося
в различных широтных поясах**

Широтный пояс, град.	Северное полушарие	Южное полушарие
0–10	14,5	15,1
10–20	14,7	14,4
20–30	14,1	15,2
30–40	15,0	15,4
40–50	13,9	14,3
50–60	11,8	11,8
60–70	8,7	7,7
70–80	5,6	4,7
80–90	1,8	1,5

ротах ОСО, в течение года претерпевающее значительные изменения (рис. 5.15), наиболее велико весной, наименее — осенью.

Озон распределен в атмосфере неравномерно по высоте. Максимальное количество его сосредоточено в стратосфере на высоте 26–27 км в тропиках, на высоте 20–21 км в средних широтах и на высоте 15–17 км в полярных областях (рис. 5.16). Как вертикальное распределение озона, так и ОСО в целом определяются равновесной концентрацией озона

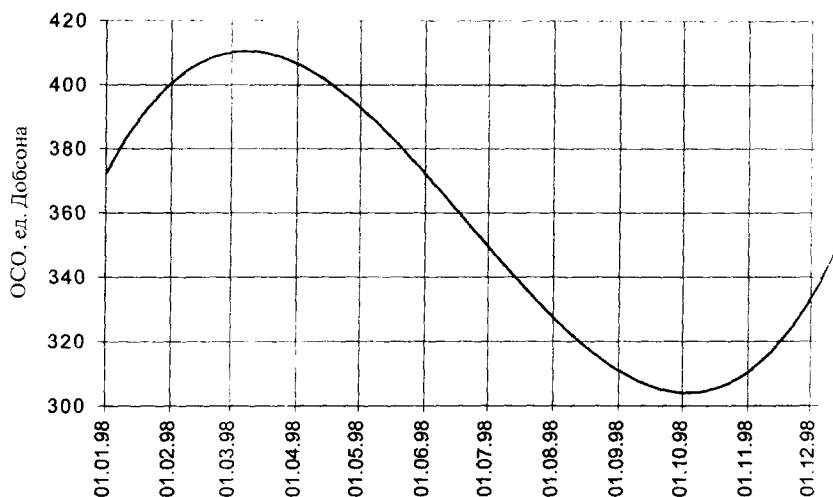


Рис. 5.15. Сглаженные ежедневные значения ОСО над Красноярском по данным спутника TOMS-EP за 1998 г.

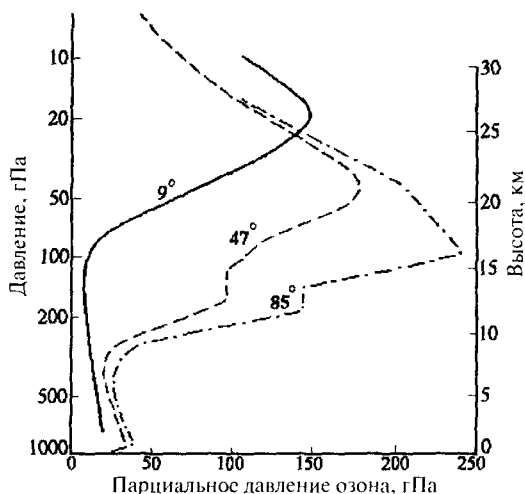


Рис. 5.16. Среднее вертикальное распределение содержания озона на различных широтах

На каждом слое атмосферы и зависят от соотношения интенсивностей процессов образования и разрушения. Вертикальные профили озона измеряют спутниковыми и наземными методами [22, 23].

Анализ полей ОСО, получаемых с помощью искусственных спутников, позволяет детально исследовать глобальные и региональные изменения озонового слоя, проследить за движениями масс озона. В качестве примера далее приведены результаты исследования пространственной динамики поля озона в области антарктической озоновой дыры.

5.5.6. Циркумполярные вихри и озоновая дыра

Одним из интереснейших природных феноменов является так называемая антарктическая озоновая дыра. Как известно, озоновая дыра — это устойчивое понижение ОСО на большой территории ниже климатической нормы.

Рекордная озоновая дыра наблюдалась в 1998 г. в Антарктике, где 30 сентября вблизи Южного полюса ОСО достигло 90 е.д. (рис. 5.17).

Образование озоновой дыры связывают с антарктическим циркумполярным вихрем — устойчивым циклоном в нижней стратосфере, который висит всю антарктическую зиму и весну над полярной областью. Внутри этого вихря, не выходя за его границы, воздух движется в основном по замкнутым траекториям вокруг Южного полюса. Поэтому в Ан-

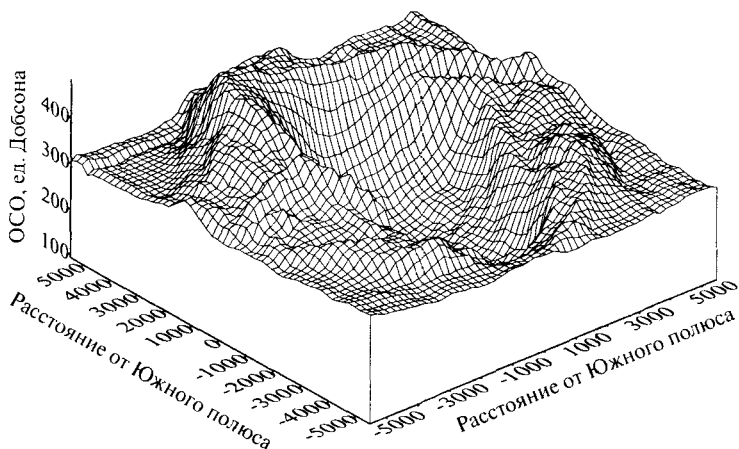


Рис. 5.17. Антарктическая озоновая дыра 30 сентября 1998 г.

тарктике зимой маловероятен обмен воздухом между полярной и среднеширотной стратосферой. Циркумполярный вихрь наблюдается весной и в Северном полушарии, но там обмен воздухом между полярной и среднеширотной стратосферой существует.

За долгую зимнюю ночь стратосферный воздух внутри антарктического вихря сильно охлаждается. Именно в конце зимы в Антарктике наблюдаются самые низкие стратосферные температуры, достигающие -70°C . Весной (сентябрь—октябрь) по мере подъема Солнца над горизонтом воздух в антарктической стратосфере начинает прогреваться, происходят повышение температуры и нарушение замкнутой картины стратосферных движений вокруг полюса (перестройка циркуляции).

Считается, что замкнутый характер движения воздушных масс в антарктической стратосфере зимой и в начале весны создает одну из предпосылок для образования дыры. Если внутри полярного вихря начнется даже относительно слабый процесс уничтожения озона, то при отсутствии обмена воздухом с другими широтными зонами к началу весны количество озона может упасть очень значительно. Полагают, что озоновые дыры не наблюдались до 80-х годов XX в. по тому, что к тому времени в стратосфере еще не накопилось достаточного количества озоноразрушающих веществ, в частности хлорсодержащих соединений.

В полярной стратосфере Южного полушария зимой (в июне—июле) на высотах 15–22 км образуются полярные стратосферные облака

енижающиеся в августе—сентябре до 10–18 км и исчезающие после потепления полярной стратосферы. Полагают, что эти облака, состоящие из кристаллов льда и капель пересохлажденной жидкости, аккумулируют активные соединения азота и тем самым создают благоприятные условия для действия хлорного цикла, разрушающего озон.

Кристаллы азотистого льда по мере аккумуляции все новых и новых молекул соединений азота становятся крупнее и тяжелее. Под действием сил притяжения они опускаются вниз в тропосферу, обеспечивая таким образом отток активных азотных соединений из области Озоновой дыры, что ведет к усилению роли хлорного цикла и разрушению озона.

Когда наступает весна, Солнце поднимается выше над горизонтом, антарктическая стратосфера постепенно прогревается. Под действием тепла и солнечных лучей полярные облака исчезают, и освободившиеся при таянии и испарении кристаллов соединения азота вновь начинают играть свою роль. К тому же идет перенос озона в область «дыры» с околполярных широт. Озоновая дыра исчезает.

Динамические процессы, обуславливающие образование вихря, наступление перестройки циркуляции и т.д., изменяются год от года. Существует квазидвухлетний цикл таких изменений и потому антарктическая озоновая дыра наблюдается приблизительно раз в два года. После 1998 г. она появилась в 2000 г.

Используя спутниковые данные об ОСО, можно оценить скорость вращения циркумполярного вихря, по крайней мере в области максимума высотной концентрации озона. Здесь озон выступает как индикатор движений в стратосфере. Можно вычислить коэффициент корреляции между полем ОСО за какой-то день и смещенным и повернутым на некоторый угол полем за предыдущий день, найти угол, соответствующий максимуму коэффициента корреляции, т.е. использовать корреляционно-экстремальный алгоритм (см § 4.4). Это позволяет определять направление и скорость движения. Наибольшее значение коэффициента корреляции, достигающее 0,7–0,95, соответствует среднему за сутки смещению и повороту поля.

Максимальная скорость вращения наблюдается в области кольца. На рис. 5.18 можно видеть, как меняется средняя за сутки угловая скорость вращения на широте 55° в Южном полушарии в сентябре 1998 г. по результатам обработки информации со спутника TOMS-EP. Здесь же для сравнения приведены аналогичные данные для Северного полушария за март того же года.

В Северном полушарии вихрь вращается с запада на восток, в Южном — с востока на запад, причем в Южном полушарии скорость вращения приблизительно в 2 раза быстрее. Вращение неравномерное, скорость

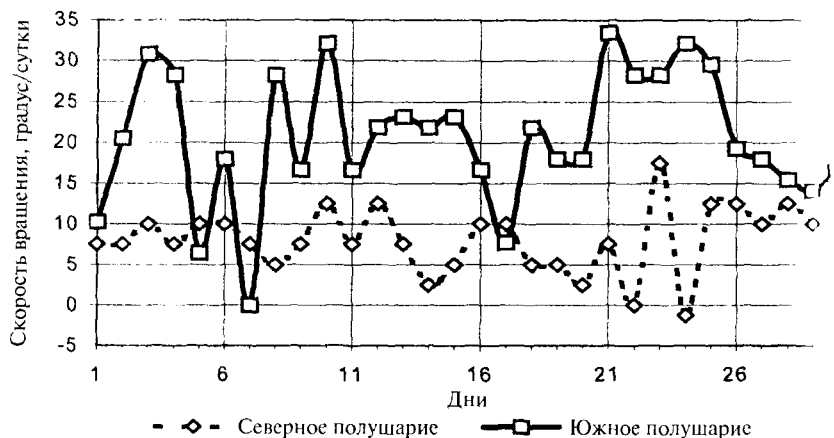


Рис. 5.18. Изменение скорости вращения циркумполярного вихря со временем

меняется изо дня в день, иногда меняется и направление. Как видно из рис. 5.19, 1–2 сентября, когда поле повернулось на 20° за сутки, коэффициент корреляции достиг 0,96. Средняя за 6 дней скорость составила только 6° в сутки из-за неравномерности вращения, а максимальное значение коэффициента корреляции 0,52. Значение коэффициента корреляции между полем за некоторый день этого периода и полем 8–10 дней позже в максимуме уменьшается до 0,2–0,3; это значит, что 8–10 дней структура стратосферного вихря существенно меняется.

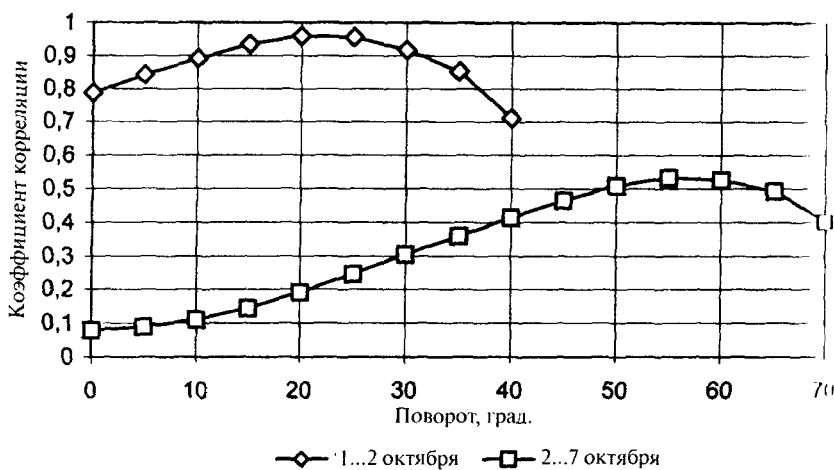


Рис. 5.19. Зависимость коэффициента корреляции от угла поворота

Именно с использованием, накоплением и анализом спутниковых данных связаны перспективы решения проблемы озоновой дыры, в частности выяснение роли естественных и антропогенных (загрязнение атмосферы) факторов в ее образовании, численного соотношения между этими двумя факторами.

6.9. ПАКЕТ IDRISI ДЛЯ ЦИФРОВОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

Для обработки данных спутникового дистанционного зондирования Земли и геоинформационных технологий удобен пакет прикладных программ IDRISI, разработанный в университете Кларк штата Массачусетс (США), имеющий русифицированный вариант. Этот пакет, получивший некоторое распространение в России, может использоваться не только в учебных, но и в научно-практических целях. Особенностью структуры пакета является включение в его состав библиотеки изображений земной поверхности, полученных с помощью сканера AVHRR спутника NOAA и сканеров TM, MSS спутника LANDSAT.

Пакет программ IDRISI объединяет порядка 100 программных модулей. Основные из них:

- работа с файлами (описание файла, редактирование файла документа, создание копии файла, переименование файла, удаление файла, импорт/экспорт);
- визуализация (вывод на экран, работа с символьными файлами, работа с файлами палитры, вывод гистограммы);
- изменение формата данных (конвертирование файлов, изменение картографической проекции, изменение масштаба, преобразование растровых изображений в векторные и наоборот и др.);
- цифровая обработка изображений (восстановление и улучшение изображений, математические операции над изображениями, вычисление статистических характеристик, распознавание образов, анализ временных серий, работа с базой данных).

Картографическая и спутниковая информация в пакете представляется в растровом и векторном виде. Растровые и векторные файлы сопровождаются файлами документации, в которых хранятся основные данные о файлах изображений, в том числе о системе географических координат, единице измерения, максимальных и минимальных координатах и др. Простота форматов растровых и векторных файлов позволяет легко создавать дополнительные опделенные пользователем модули.

IDRISI снабжен системой, которая позволяет импортировать и экспортировать графические файлы в форматах TIFF, BMP, DXF и файлы в форматах, применяемых в ГИС-системах: ARCINFO, ATLAS*GIS BNA, ERDAS, GRASS, Map Analysis Package (MAP), MAPINFO и др.

Процедуры улучшения изображений предусматривают гистограммные преобразования: эквализацию и растяжку гистограмм. Многоспектральные спутниковые изображения могут быть представлены в условных цветах. Для цифровой фильтрации в пространственной области используются линейные

фильтры — сглаживающие и градиентные; нелинейные фильтры представлены медианным фильтром и фильтром Собела. Размеры окон фильтров различные — от 3×3 до 7×7 . У пользователя существует также возможность самому задавать весовые множители фильтров.

Классификация изображений предусматривает применение байесовской процедуры, метода максимального правдоподобия, метода параллелепипеда, метода минимальных расстояний и др. Возможны вычисление вегетационного индекса и анализ главных компонент, кластерный анализ.

Для статистического анализа IDRISI имеет ряд модулей, которые позволяют производить расчет основных статистических характеристик, в том числе функции автокорреляции, линейной регрессии между двумя изображениями и др.

В руководстве к пакету IDRISI приводятся краткие теоретические сведения о ГИС, необходимые сведения из геодезии и картографии, введение в дистанционное зондирование и цифровую обработку изображений.

Пакет содержит удачно сформированные методически упражнения по ГИС, по преобразованию и цифровой обработке изображений, методам распознавания образов и многое другое.

Последняя версия пакета IDRISI-32 ориентирована на современные компьютеры, обладает расширенными возможностями, в частности предусмотрено использование процедуры Криге, двумерного преобразования Фурье, введены новые процедуры распознавания образов, предусмотрена обработка изображений, получаемых с помощью радиолокаторов бокового обзора с синтезированной апертурой, и др.

IDRISI используется более чем в 130 странах при обучении студентов при обработке аэрокосмических изображений и как пакет ГИС.

С IDRISI можно познакомиться в сети Интернет по адресу <http://www.idrisi.clarku.edu>.

КОНТРОЛЬНЫЕ ЗАДАНИЯ

1. Что такое ГИС? Каковы ее основные задачи?
2. На основании каких данных производится долгосрочный, среднесрочный и краткосрочный прогнозы пожарной опасности в лесах?
3. Как определяется начало пожароопасного сезона в лесу?
4. Что такое ресурсная облачность?
5. Как организована база данных по уровням ГИС?
6. Какова схема возникновения и развития лесного пожара? Каковы физические характеристики лесного пожара?
7. На какие длины волн приходится максимум излучения от лесного пожара, какой спутник вы выберите для регистрации восходящего излучения на этих длинах волн?
8. Всегда ли лесные пожары приносят только вред?
9. Как с помощью искусственного спутника Земли оценить пожарную опасность в лесу?
10. Как с помощью искусственного спутника Земли обнаружить лесной пожар? Как отличить пожар от ложных засветок?

11. Как по данным сканера определить площадь пожара, если он занимает часть пиксела?
12. Как использовать вегетационный индекс для контроля состояния растительности? Можно ли по вегетационному индексу оценить урожай на поле?
13. По рис. 5.5 и 5.6 оцените степень повреждения лесной растительности в районе Норильска в период с 1979 по 1984 г. Каков процент техногенного повреждения леса?
14. Какую роль играют фронтальные зоны в океане? Как их обнаружить по спутниковым данным?
15. Сформулируйте проблемы, с которыми приходится сталкиваться при оценке площади заснеженности по спутниковым данным.
16. Как использовать вегетационный индекс для разделения участков на изображениях, отвечающих облакам, снегу/льду, растительному покрову?
17. Каковы преимущества радиолокационного обзора льда перед обзором его в оптическом диапазоне?
18. Опишите физические принципы, лежащие в основе определения вертикальных профилей температуры в атмосфере с искусственных спутников Земли, преимущества и недостатки дистанционного определения профилей температуры и влажности.
19. Каковы основные метеорологические особенности теплого и холодного атмосферных фронтов и как это отражается на космических изображениях?
20. По каким признакам на космическом изображении облачности можно прогнозировать возникновение шквального ветра?
21. В чем состоит экологическая роль атмосферного озона?
22. Выведите формулу для расчета эффективного значения ОСО в полдень с учетом широты местности и времени года.
23. В чем состоят основные химические механизмы разрушения озонового слоя?
24. Сформулируйте принципы определения ОСО космическими методами в оптическом диапазоне.
25. Используя рис. 5.14, оцените общее содержание озона над Новосибирском, Красноярском, Иркутском и Норильском, минимальное и максимальное ОСО в регионе.
26. Как с помощью корреляционно-экстремального алгоритма оценить направление и скорость смещения поля ОСО?

ЛИТЕРАТУРА

1. Геоинформационные системы. Обзорная информация. М.: ЦНИИГАИК, 1992. С. 52.
2. Аэрокосмический мониторинг лесов / А.С. Исаев, В.И. Сухих, Е.Н. Калашников и др. М.: Наука, 1991.
3. Даниюлис Е.П., Жирин В.М., Сухих В.И., Эльман Р.И. Дистанционное зондирование в лесном хозяйстве. М.: Агропромиздат, 1989.

4. Сухинин А.И. Система космического мониторинга лесных пожаров в Красноярском крае // Сибирский экологический журнал. 1996. Т. 3, № 1. С. 85–92.
5. Валендик Э.Н., Сухинин А.И., Кисильхов Е.К., Хребтов Б.А. Мониторинг лесных пожаров // Исследование лесов аэрокосмическими методами. Новосибирск: Наука, 1987. С. 118–135.
6. Sukhinin A.I., Kashkin V.B., Ponomarev E.I. Monitoring Forest Fires in Eastern Siberia from Space // Proc. of SPIE. V. 3983. P. 206–214.
7. Ромасько В.И., Кашкин В.Б., Сухинин А.И. Послепожарная инвентаризация леса по спутниковым данным // Исследование Земли из космоса. 1993. № 6. С. 99–103.
8. Coward S.N. et al. Normalized Difference Vegetation Index Measurement from the Advanced Very High Resolution Radiometer // Remote Sensing Environment. 1991. V. 35, № 2–3. P. 237–278.
9. Wiegand C.L. et al. Vegetation Indices in Crop Assessments // Remote Sensing Environment. 1991. V. 35, № 2–3. P. 105–120.
10. Шевырногов А.П., Сидько А.Ф. Спектральная яркость растений как основа дистанционной диагностики посевов сельскохозяйственных культур / ДАН. 1997. Т. 354, № 1. С. 11–118.
11. Оценка данных AVHRR для прогноза урожайности зерновых культур П.Р. Попович, А.А. Феоктистов, Е.Л. Злобин и др. // Исследование Земли из космоса. 1993. № 1. С. 99–102.
12. Кондратьев К.Я., Вандышева Н.В., Козодеров В.В., Косолапов В.С. Оценка параметров почвенно-растительного покрова по многоспектральным спутниковым данным // Исследование Земли из космоса. 1992. № 3. С. 88–95.
13. Трахов Э.М. и др. Прогнозирование урожайности озимой пшеницы с использованием материалов аэрофотосъемки // Исследование Земли из космоса. 1992. № 5. С. 102–106.
14. Харук В.И., Винтербергер К., Цибульский Г.М., Яхимович А.П. Анализ техногенной деградации притундровых лесов по данным съемки из космоса // Исследование Земли из космоса. 1995. № 4. С. 91–97.
15. Kartushinsky A.V. Time-Space Structure and Variability of Surface Temperature Frontal Zones in the Ocean (Based on AVHRR Satellite Data) // Adv. Space Res. 2000. V. 25, № 5. P. 1107–1110.
16. Методика определения заснеженности речного бассейна по спутниковым данным для оперативных прогнозов стока / Д.А. Бураков, В.Б. Кашкин, А.И. Сухинин и др. // Метеорология и гидрология. 1996. № 8. С. 100–109.
17. Бураков Д.А., Авдеева Ю.В. Технология оперативных прогнозов ежедневных расходов (уровней) воды на основе спутниковой информации о заснеженности (на примере р. Нижняя Тунгуска) // Метеорология и гидрология. 1996. № 10. С. 75–87.
18. Радиолокационные методы исследования Земли / Под ред. Ю.А. Мельника. М.: Сов. радио, 1980. 262 с.
19. Кондратьев К.Я., Тимофеев Ю.М. Метеорологическое зондирование атмосферы из космоса. Л.: Гидрометеиздат, 1978. 279 с.
20. Герман М.А. Метеорологическое зондирование атмосферы из космоса. Л.: Гидрометеиздат, 1985. 260 с.

21. *Матвеев Л. Т.* Курс общей метеорологии. Физика атмосферы. Л.: Гидрометеониздат, 1984. 750 с.
22. *Хргиан А. Х.* Физика атмосферы. М.: Изд. МГУ, 1986. 328 с.
23. *Александров А. Л., Израэль Ю. А., Кароль И. Л., Хргиан А. Х.* Озоновый щит Земли и его изменения. СПб: Гидрометеониздат, 1992. 287 с.
24. *Кашкин В. Б., Ромасько В. И., Сальникова О. Е., Сухинин А. И.* Озоновый слой в Сибири по данным спутника NOAA в ИК-диапазоне летом 1995 г. // Исследование Земли из космоса. 1997. № 5. С. 93–96.
25. *Kashkin V. B., Gomonov A. A., Sukhinin A. I.* Researches of the Ozone Layer using Satellite Data // Proc. of SPIE. V. 3983. P. 538–541.
26. *Кашкин В. Б.* Влияние военных операций НАТО на Балканах на озоновый слой Земли // Тезисы докладов VI Международного симпозиума «Оптика атмосферы и океана». Томск: ИОА СО РАН, 1999. С. 87.

Учебное издание

Кашкин Валентин Борисович
Сухинин Анатолий Иванович

**ДИСТАНЦИОННОЕ ЗОНДИРОВАНИЕ
ЗЕМЛИ ИЗ КОСМОСА
ЦИФРОВАЯ ОБРАБОТКА ИЗОБРАЖЕНИЙ**

Учебное пособие

Редактор Э.М. Горелик
Переплет Е. Молчанова, С. Носова
Компьютерная верстка П. Аборина
Корректор Т.Г. Тертышная

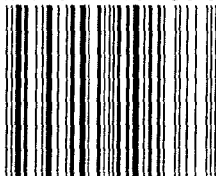
Изд. лиц. ИД № 01670 от 24.04.2000
Налоговая льгота – общероссийский классификатор
продукции ОК-005-93, том 2: 953000

Подписано в печать 10.12.2001. Формат 60х90/16.
Печать офсетная. Бумага офсетная. Печ. л. 16,5.
Гарнитура Школьная. Тираж 1000 экз. Заказ № 2500.

Издательско-книготорговый дом «Логос»
105318, Москва, Измайловское ш., 4

Отпечатано в РГУП «Чебоксарская типография №1»
428019, Чебоксары, пр. И.Яковлева

ISBN 5-94010-138-0



9 785940 101383