
В. В. СЕЛЯНКИН



КОМПЬЮТЕРНОЕ ЗРЕНИЕ АНАЛИЗ И ОБРАБОТКА ИЗОБРАЖЕНИЙ



Учебное пособие

Издание второе, стереотипное



ЛАНЬ

• САНКТ-ПЕТЕРБУРГ •
• МОСКВА • КРАСНОДАР •

• 2021 •

УДК 004.932
ББК 32.81я73

С 29 **Селянкин В. В.** Компьютерное зрение. Анализ и обработка изображений : учебное пособие для вузов / В. В. Селянкин. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 152 с. : ил. — Текст : непосредственный.

ISBN 978-5-8114-8259-7

В учебном пособии рассматриваются вопросы анализа, синтеза и обработки полутоновых и цветных изображений, сегментации изображений. Описанные методы работы с изображениями используются в решении многих задач компьютерного зрения: распознавания образов, поиск изображений на основе содержания, обнаружение движения на изображениях, сопоставление изображений в двумерном пространстве, распознавание объектов на изображениях и другие.

Пособие рассчитано на студентов магистерской подготовки, аспирантов и специалистов, работающих в области использования компьютерной графики и компьютерного зрения.

УДК 004.932
ББК 32.81я73



Обложка
Ю. В. ГРИГОРЬЕВА

© Издательство «Лань», 2021
© В. В. Селянкин, 2021
© Издательство «Лань»,
художественное оформление, 2021

СОДЕРЖАНИЕ

Введение	7
1. Средства и методы формирования изображений	12
1.1. Устройства для формирования изображений	12
1.1.1. Камеры на основе приборов с зарядовой связью	12
1.1.2. Видеокамеры	12
1.1.3. Микроденситометры	13
1.1.4. Датчики для цветных и мультиспектральных изображений ...	13
1.1.5. Датчики рентгеновского изображения	14
1.1.6. Изображения, полученные методом ЯМР	14
1.1.7. Дальнометрические изображения	15
1.1.8. Ультрасонографические изображения	15
1.2. Особенности формирования цифровых изображений	15
1.3. Системы координат для трехмерных изображений	16
1.3.1. Экранная система координат	17
1.3.2. Система координат объекта	17
1.3.3. Система координат камеры	17
1.3.4. Декартова система координат изображения	17
1.3.5. Мировая система координат	18
2. Анализ бинарных изображений	19
2.1. Обработка черно-белых изображений	19
2.2. Подсчет объектов на изображении	20
2.3. Маркировка связных компонент изображения	22
2.4. Морфология бинарных изображений	23
2.4.1. Структурирующие элементы	23
2.4.2. Основные операции бинарной морфологии	24
2.5. Свойства бинарных областей	25
2.5.1. Площадь области и центр тяжести	26
2.5.2. Длина периметра	26
2.5.3. Свойство округлости	27
2.5.4. Описывающий прямоугольник и экстремальные точки	28
2.5.5. Моменты второго порядка для областей	29
2.6. Графы смежности областей	31
2.7. Пороговая бинаризация изображений	32
3. Обработка полутоновых изображений	34
3.1. Увеличение диапазона полутонов	34
3.2. Удаление мелких деталей	36
3.2.1. Удаление шума «соль и перец»	36
3.2.2. Удаление мелких областей	36

3.2.3. Сглаживание изображений.....	37
3.2.4. Медианный фильтр.....	37
3.3. Обнаружение краев с помощью дифференциальных масок.....	38
3.3.1. Дифференцирование одномерных сигналов.....	38
3.3.2. Дифференцирование двумерных изображений.....	41
3.3.3. Обнаружение краев с помощью фильтрации.....	43
3.3.4. Обнаружение краев с помощью ЛОГ-фильтра.....	44
3.4. Использование масок для анализа изображения.....	45
3.4.1. Некоторые понятия векторов.....	45
3.4.2. Анализ на основе базиса Робертса.....	47
3.4.3. Анализ на основе базиса Фрея – Чена.....	48
4. Обработка цветных изображений.....	52
4.1. Физические свойства цвета.....	52
4.2. Цветовые модели.....	53
4.2.1. Цветовая модель RGB.....	53
4.2.2. Цветовая модель CMY.....	53
4.2.3. Цветовая модель HIS.....	54
4.2.4. Цветовые системы YIQ и YUV.....	54
4.3. Гистограммы цветных изображений.....	54
4.4. Текстура.....	55
4.4.1. Понятие текстуры.....	55
4.4.2. Тексельное описание текстур.....	56
4.4.3. Количественные характеристики текстур.....	57
4.4.4. Текстуальная сегментация.....	60
5. Сегментация изображений.....	62
5.1. Понятие задачи сегментации.....	62
5.2. Обнаружение областей.....	62
5.3. Методы кластеризации.....	63
5.3.1. Классические алгоритмы кластеризации.....	63
5.3.2. Итерационная кластеризация по математическому ожиданию.....	63
5.3.3. Алгоритм кластеризации ISODATA.....	64
5.3.4. Простые гистограммные методы.....	65
5.3.5. Рекурсивный гистограммный метод.....	66
5.3.6. Алгоритм графового разбиения Ши.....	66
5.4. Нарращивание областей.....	68
5.5. Способы представления областей.....	69
5.5.1. Оверлейные представления.....	69
5.5.2. Маркированные изображения.....	69
5.5.3. Кодирование границ.....	69

5.5.4. Квадрантные деревья	70
5.5.5. Таблицы свойств.....	71
5.6. Обнаружение контуров	71
5.6.1. Прослеживание границ областей	71
5.6.2. Обнаружение краев Кэнни.....	72
5.6.3. Группировка контурных фрагментов	72
5.6.4. Обнаружение прямых и дуг окружностей.....	73
5.6.5. Обнаружение отрезков прямых и окружностей	73
5.7. Подбор моделей сегментов	74
5.7.1. Подбор прямой	75
5.7.2. Подбор других моделей сегментов	76
6. Методы распознавания образов.....	77
6.1. Основные задачи распознавания образов.....	77
6.2. Ошибки распознавания	78
6.3. Распознавание по набору измерений	80
6.4. Распознавание по элементарным признакам	80
6.4.1. Нахождение класса объекта по среднему значению	80
6.4.2. Нахождение класса по расстоянию до ближайших соседей ...	82
6.5. Структурные методы распознавания	82
6.6. Матрица неточностей	83
6.7. Дерево решений	84
6.8. Автоматическое построение дерева решений	87
6.9. Байесовский метод принятия решений.....	89
6.10. Понятие методов кластеризации	90
6.11. Искусственные нейронные сети	91
7. Поиск изображений на основе содержания	93
7.1. Задача поиска изображений.....	93
7.2. Поиск с использованием ключевых слов	93
7.3. Поиск с использованием запросов по образцу.....	94
7.4. Меры расстояния между изображениями.....	95
7.4.1. Характеристики цветового сходства	95
7.4.2. Характеристики текстурного сходства.....	96
7.4.3. Характеристики сходства формы.....	98
7.4.4. Характеристики сходства объектов и их расположения на изображении	102
7.5. Организация базы данных.....	106
7.5.1. Стандартные индексы	106
7.5.2. Пространственная индексация	109
7.5.3. Индексация в системах поиска изображений на основе содержания	109

8. Обнаружение движения на изображениях.....	111
8.1. Задача обнаружения движения.....	111
8.2. Вычитание изображений.....	111
8.3. Вычисление векторов перемещения.....	112
8.4. Вычисление траекторий движущихся точек.....	114
8.5. Виды изменений условий съемки.....	116
8.6. Сегментация видеопоследовательностей.....	117
9. Сопоставление изображений в двумерном пространстве.....	119
9.1. Основные понятия.....	119
9.2. Распознавание двумерных объектов с помощью аффинных преобразований.....	120
9.3. Распознавание двумерных объектов с использованием реляционных моделей.....	124
10. Распознавание объектов на изображениях.....	129
10.1. Определение соответствия точек двумерного изображения и точек трехмерной сцены.....	129
10.2. Распознавание объектов на основе моделей.....	129
10.3. Существующие разновидности трехмерных моделей.....	130
10.4. Классы видимости трехмерных объектов.....	135
10.5. Физические и деформационные модели.....	135
10.6. Методы распознавания трехмерных объектов.....	137
10.6.1. Классификация задач распознавания.....	137
10.6.2. Распознавание с использованием процедуры совмещения.....	139
10.6.3. Распознавание на основе реляционных моделей.....	142
Заключение.....	147
Библиографический список.....	148



Введение

Термин «компьютерное зрение» в современной технической литературе соседствует с похожими понятиями: «техническое зрение» и «машинное зрение». Во многих случаях их можно воспринимать как синонимы, хотя возможны и отличия в некоторых практических приложениях. Например, «машинное зрение» может применяться в контексте промышленных приложений. Для этого случая в русскоязычной литературе в качестве синонима используется и термин «техническое зрение» [1]. В данном учебном пособии термины «компьютерное зрение» и «машинное зрение» считаются синонимами.

Кроме того, можно отметить, что компьютерное зрение базируется на элементах компьютерной графики и является ее развитием в практическом использовании в различных областях знания [1–4]. В качестве таких областей можно указать следящие и охранные системы, системы контроля качества изделий, распознавание объектов и текстовых документов, поиск изображений на основе содержания, медицинскую диагностику, получение и обработку спутниковых изображений, робототехнику, контроль на транспортных магистралях и другие. Здесь можно говорить о некоторых аспектах зрения человека, реализованных техническими средствами. Хотя качество решения этих задач несравнимо с человеком, но даже такой уровень решения задач является существенным прогрессом современных IT-технологий.

Последовательность операций для получения конечного результата решения задачи компьютерного зрения может иметь вид, показанный на рисунке 1.



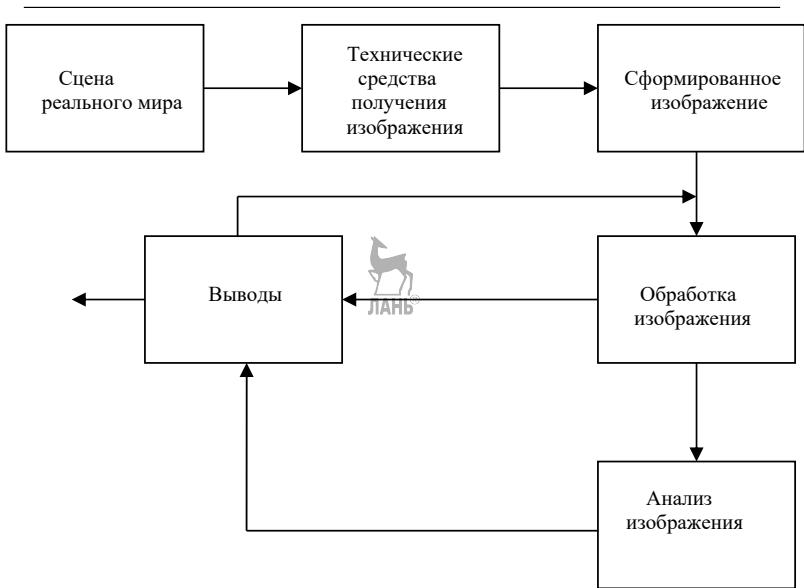


Рис. 1. Схема конвейера решения задачи компьютерного зрения

В зависимости от вида решаемой задачи конкретизируется наполнение каждого блока представленного конвейера. Ниже приведены примеры реализации такого конвейера для различных практических задач.

Задача поиска изображений на основе содержания. Имеется база данных изображений. Такими изображениями могут быть лица людей, различные объекты трехмерного мира, изображения автомобилей, товаров, различных изделий, зданий и т. п. Ставится задача поиска нужного изображения в базе данных. Причем задача может формулироваться для поиска как точного изображения, так и максимально приближенного к заданному изображению. Возможно также выделение класса объектов, к которому принадлежит заданный объект. В этом случае в качестве сцены реального мира выступают соответствующие объекты и для получения базы данных изображений используются необходимые технические средства.

Важной задачей является подбор методов и алгоритмов обработки изображений, которые могут дать качественное решение задачи. Технологии обработки в этом случае зависят от многих параметров: объема базы данных, размеров на изображении, качества изображений, парамет-

ров яркости и контрастности изображений, наличия фона, ракурсы расположения объектов и т. д.

В зависимости от требований задачи возможен широкий спектр требуемых выводов. В простейших случаях возможен двузначный ответ типа «да/нет», в более сложных – выбор нужного объекта на изображении или класса объектов.

Задача обработки медицинских изображений. В качестве изображений используются изображения внутренних органов человека. Для их получения используют эффект ядерного магнитного резонанса для головы, рентгенографию для изображений легких и скелета, компьютерную томографию для различных органов, ультразвукографию для сердца, почек и других органов. Диагностика состояний внутренних органов базируется на привлечении современных медицинских знаний и выработки качественных методов анализа изображений.

Задача обработки текстовых документов. Здесь объектом реального мира выступает текстовый документ, по которому получают цифровую форму его изображения с помощью сканера. Задача компьютерного зрения может формулироваться как распознавание отдельных символов, слов или перевода цифрового изображения в текстовый файл. Кроме того, может ставиться задача и семантического распознавания текста для индексации его в большой базе данных.

Задача сортировки деталей на конвейере. В таких задачах предполагается ограниченный набор деталей, что позволяет сформировать для них вектор характерных параметров. Полученные изображения деталей с помощью камеры (или нескольких камер) необходимо обработать в реальном масштабе времени и выполнить сортировку деталей. При распознавании детали может использоваться трехмерная модель, предварительно сформированная с помощью их характерных параметров.

Задача обработки видеопотока для охранных систем. Как правило, такие задачи связаны с распознаванием движения объектов на фоне неподвижной сцены. Здесь также задействованы камеры, настроенные на передачу изображения некоторой сцены. Априорно предполагается, что сцена в процессе времени остается неизменной или почти неизменной, т. е. динамика движения возможна в ограниченных пределах. Задача заключается в обнаружении движущихся объектов, возможно без распознавания самих объектов. Похожая задача может выполняться и для различных следящих систем.

Анализ транспортной магистрали. Данная задача близка к предыдущей и отличается от нее некоторыми деталями по скорости движения объектов и возможными элементами распознавания типа движущего

объекта и номера автомобиля. При этом может ставиться задача фиксации дорожно-транспортных происшествий.

Задача обработки спутниковых изображений. Обработка спутниковых изображений может иметь разные аспекты. Могут решаться задачи сельскохозяйственного назначения, анализа и состояния лесных массивов, обнаружения пожаров, оценки состояния снежного покрова перед весенними паводками, задачи распознавания военного назначения и другие. На решение этих задач оказывают существенное значение качество аппаратуры для съемок, погодные условия и скорость обработки информации. Во многих случаях для эффективного решения задач может быть полезным использование мультиспектральных изображений.

В задачах компьютерного зрения существенное значение имеет обработка изображений для получения необходимой информации. Несмотря на существование большого количества приемов такой обработки, их можно разбить на несколько категорий в зависимости от назначения и методики обработки. Например, можно рассмотреть такие категории:

- формирование изображения по заданному описанию;
- повышение качества изображения;
- обработка изображения для получения изображения с заданными свойствами;
- анализ изображения для получения результата описательного характера.

Рассмотрим более подробно содержание этих методов [1].

Корректировка окрестностей пикселей. Назначение этих методов состоит в подавлении шумов и упрощении изображения за счет удаления мелких деталей изображения. Иногда применяется корректировка граничных пикселей сплошных областей изображения. Эта операция в некоторых случаях позволяет четче разделить отдельные элементы изображения. Корректировка граничных пикселей заключается в замене соседних пикселей на фоновые значения.

Улучшение качества изображения. Операция выполняется для всего изображения по одному алгоритму. Улучшение достигается за счет изменения яркости (повышения или понижения), улучшение гистограммы изображений, сглаживания, изменения контрастности, а также за счет удаления шумов.

Получение изображения по нескольким исходным. В этом случае используются операции вычитания или сложения изображений. Операция вычитания позволяет выявить движущиеся объекты, а сложения – получить новое изображение с суммарными элементами исходных изображений.

ражений. Как правило, в качестве исходных используются два изображения, но возможны и другие варианты.

Вычисление характеристик изображения. В зависимости от решаемой задачи можно подобрать характерные признаки распознаваемых объектов: длина, ширина, площадь, цвет, характерные контуры, ориентация и прочие. По составленному набору характеристик проводится подсчет их на изображении. Установив некоторые допустимые диапазоны значений заданных характеристик, можно проводить анализ результатов вычислений и делать итоговые выводы.

Получение описательного результата. Эти методы относятся большей частью именно к задачам компьютерного зрения. Результаты решения задачи могут выдаваться в виде подсчета некоторых объектов на изображении, наличия или отсутствия каких-то признаков, в виде текстового файла, полученного по его графическому изображению. Возможно получение описания сцены с указанием перечня выявленных объектов, их характеристиками и ориентацией на сцене.

Задачи компьютерного зрения в настоящее время широко востребованы во многих областях деятельности человека. Это объясняется успехами современных IT-технологий в области программно-аппаратных средств вычислительной техники – появление мощных центральных и графических процессоров, языковых и алгоритмических средств. Однако, как уже говорилось, результаты решения задач компьютерного зрения сегодня значительно уступают зрению человека и требуют дальнейшего совершенствования.



1. Средства и методы формирования изображений

1.1. Устройства для формирования изображений

1.1.1. Камеры на основе приборов с зарядовой связью

Такие камеры создаются с использованием приборов с зарядовой связью (ПЗС). Свет от внешнего источника (трехмерная сцена) через объектив попадает на плоскость изображения в виде дискретной матрицы, состоящей из твердотельных ячеек, преобразующих световую энергию в электрический заряд. Предварительно очищенные ячейки накапливают заряд в течение некоторого времени, определяемого временем открытия затвора. Матрица ячеек обладает памятью, которая затем считывается и, после аналого-цифрового преобразователя, передается в буфер кадра. Далее это изображение может передаваться в компьютер для дальнейшей обработки.

1.1.2. Видеокамеры

Видеокамеры предназначены для формирования последовательности кадров, отображающих трехмерную сцену в динамике. Кадры формируются с определенной частотой, обычно 30 кадров в секунду. Для повышения непрерывности восприятия движения человеком используют прием представления кадра двумя блоками с прореженным через строчку изображением. Такой метод носит название чересстрочной развертки кадра. Блоки в этом случае содержат четные и нечетные строки кадра. Частота выдачи кадров при этом составляет 60 раз в секунду, что для человека дает лучшее восприятие изображения за счет свойства инерционности глаза. Современные видеокамеры, ориентированные на машинную обработку, обладают достаточной скоростью воспроизведения, поэтому развертка кадра выполняется построчно без разбиения на полукадры. Следует отметить, что видеокамеры кроме изображения записывают и звуковое сопровождение, а сами кадры разделяются маркерами.

Для уменьшения объема информации видеопоследовательности записываются с использованием каких-то способов сжатия. Для чего существуют различные форматы записи и сжатия.

1.1.3. Микроденситометры

Датчики этого типа предназначены для сканирования слайдов или фотопленок на просвет. В этом случае сканируемые кадры размещаются между источником света и датчиком. Датчик записывает значения, пропорциональные интенсивности прошедшего через кадр света. Так как датчик является точечным, то для получения полного изображения необходимо перемещать кадр для его сканирования. При этом требуется достаточно точная механика перемещения. Преимущество микроденситометров заключается в том, что у них отсутствуют ошибки измерения интенсивности по сравнению с ПЗС-матрицей. Кроме того, они дают большее разрешение изображения. Главным недостатком их является пониженная скорость обработки.

1.1.4. Датчики для цветных и мультиспектральных изображений

Цветные изображения могут получаться в ПЗС-камерах при использовании разноцветного вращающегося диска. Диск изготовлен из прозрачной пленки, окрашенной в основные цвета: красный-зеленый-синий. Сектора разных цветов имеют равные площади. Скорость вращения диска подбирается таким образом, чтобы в течение одного считывания перед датчиком проходил только один цвет. За один оборот диска производится три считывания, в результате чего выполняется формирование трех компонент цветности одного пикселя. Процесс формирования пикселя длится в течение одного оборота диска, что приводит к пространственным искажениям изображения, так как все три компонента считываются в разные моменты времени движущегося спутника.

Мультиспектральные изображения получают в таких же ПЗС-камерах, в которых используются тонкие преломляющие пленки. Преломляющие средства разделяют падающий луч на спектр из 4 компонент. Каждая компонента фиксируется своим элементом ПЗС-матрицы, что в совокупности формирует набор цветных изображений, полученных для световых волн разной длины.

На спутниках могут устанавливаться сканеры направленного зондирования земной поверхности. При этом каждый участок земной поверхности сканируется направленной оптической системой с одновременным подавлением излучения остальных участков. Излучение от сканируемого участка проходит через призму, которая разлагает его в спектр. Полученный спектр разбивается на диапазоны, каждый из которых фиксируется ПЗС-матрицей в самостоятельное изображение. По числу диапазонов для

каждого пикселя формируется вектор значений интенсивности соответствующей длины. За один проход получается одна строка двумерного изображения. В процессе движения спутника по орбите формируются последовательно все строки двумерного изображения. Данная технология получения мультиспектрального изображения также характеризуется пространственными искажениями. Мультиспектральные изображения показывают удовлетворительные результаты распознавания характера земной поверхности – водные, лесные сельскохозяйственные, горные и другие.

1.1.5. Датчики рентгеновского изображения

Источники и датчики рентгеновского излучения используются в медицине для получения изображений частей тела человека, в машиностроении – для контроля сварных и литых деталей, контроля продукции в пищевой промышленности и других областях. Такие изображения фиксируют интенсивность рентгеновских лучей, которые проходят через анализируемый объект. В зависимости от плотности и качества материала меняется распределение этой интенсивности по изображению.

Данная технология используется и для получения трехмерной картины наблюдаемого объекта. В этом случае применяются рентгеновские томографы. Полное представление картины достигается последовательным сканированием объема тела и формированием вокселей, отображающих объемный характер тела. Полученная таким образом информация о трехмерном объекте позволяет при необходимости анализировать его в нужных направлениях и плоскостях в виде двумерных изображений.

1.1.6. Изображения, полученные методом ЯМР

Другой метод получения трехмерных изображений основан на эффекте ядерного магнитного резонанса для объектов, состоящих из различных веществ. Как правило, ЯМР-томографы используются в медицине для диагностики различных заболеваний. При получении изображения томограф регистрирует интенсивность сигнала ЯМР для каждого вокселя – элемента трехмерного пространства.

В магниторезонансной ангиографии интенсивность сигнала зависит от скорости потока крови, что позволяет диагностировать многие заболевания, связанные с состоянием сердечно-сосудистой системы пациента. Методику ЯМР можно использовать и в других сферах деятельности человека, но в настоящее время расширение сфер применения сдерживается высокой стоимостью аппаратуры и процедуры сканирования.

1.1.7. Дальнометрические изображения

Существуют устройства, позволяющие получать изображения с отображением интенсивности излучения элементов трехмерных поверхностей и расстояния до них. Эти устройства-дальномеры имеют лазерный излучатель с амплитудной модуляцией. Излучатель посылает лазерный луч, затем принимает отраженный сигнал. Расстояние до отражающей поверхности вычисляется на сравнении фазы излученного и отраженного сигнала. Как правило, таким способом измеряются расстояния в пределах одного периода волны модулирующего сигнала. Для больших расстояний вычисление возможно, если известно число полных периодов сигнала.

Так как датчики позволяют измерять и интенсивность отраженного сигнала, то формируются два совмещенных изображения – изображения дальности и интенсивности в полутоновом виде. Эти устройства имеют невысокую скорость обработки, а аппаратура отличается высокой стоимостью, что сужает области их использования. В настоящее время они применяются в горнодобывающей робототехнике, в космических роботах, в автомобильной промышленности и некоторых других областях.

1.1.8. Ультрасонографические изображения

Исследования с помощью ультразвуковой аппаратуры (УЗИ) широко используют в медицине. Они носят название ультрасонографии – метода получения изображений с использованием высокочастотных звуковых волн. В медицине УЗИ используют для диагностики внутренних органов – печени, желчного и мочевого пузыря, почек, щитовидной железы и других. Для кровеносных сосудов используют ультразвуковое исследование – доплерографию, с помощью которой определяют направление и скорость потока крови, а также изменения в сосудах, вызванные болезнью, их функциональное состояние.

1.2. Особенности формирования цифровых изображений

При формировании цифровых изображений могут возникать различные **геометрические искажения**. Это обычно связано с несовершенством технических средств и точности их изготовления. Они дают искажения форм, линий, окружностей, а также других элементов изображений. Причем искажения могут проявляться неравномерно по площади изображения.

При съемках объектов, удаленных на большие расстояния, могут оказывать влияние состояния погодных условий, наличия влаги в атмосфере, температуры воздуха. Эти условия влияют на прохождение лучей, их преломление и рассеивание. Такие искажения называются **дисперсией**.

Искажения – **блюминг** (избыточная яркость) – объясняются влиянием элементов ПЗС-матрицы на соседние элементы. В этом случае часть накопленного заряда элемента ПЗС-матрицы при ярком освещении передается на соседние элементы, что приводит к увеличению размеров ярко-го пятна.

Неоднородность элементов ПЗС-матрицы за счет погрешностей технологии их изготовления приводит к разной их чувствительности к яркости света. В связи с этим элементы с одинаковой освещенностью могут давать различные показания. Снижение этого искажения добиваются частично за счет калибровки элементов ПЗС-матрицы, при которой для них вырабатываются масштабные коэффициенты, позволяющие корректировать считываемые значения яркости. Возможны случаи, когда отдельные элементы матрицы могут вообще не действовать. Таким ячейкам присваиваются значения яркости, вычисленные как средние значения соседних ячеек.

При формировании цифровых изображений следует учитывать **диапазон** допустимых значений интенсивности света. Выход за границы диапазона дадут искажения, существенно изменяющие истинные значения яркости. Чтобы исключить такие искажения, следует увеличивать диапазон возможных значений или вести контроль выхода за его пределы.

Световые волны различной длины преломляются линзой в разной степени. Поэтому световые волны от точечного элемента сцены после преломления могут попасть сразу на несколько соседних пикселей, давая при этом **эффект размытия**.

Преобразование аналогового сигнала в цифровой дает дополнительное искажение за счет **дискретизации и округления** значения сигнала.

1.3. Системы координат для трехмерных изображений

При формировании трехмерных изображений, их обработке и анализе возможно использование пяти различных систем координат. В каждом конкретном случае могут применяться часть этих систем. С другой стороны, каждый объект трехмерной сцены имеет собственную объектную

систему координат, и в этом смысле число используемых систем координат будет зависеть от состава трехмерной сцены. Ниже рассматриваются виды этих систем.

1.3.1. Экранная система координат

Эта система координат S имеет две оси X и Y . Ось X направлена слева направо, а ось Y – сверху вниз. Пределы изменения значений определяются разрешающей способностью экрана, а экран представляется целочисленной решеткой пикселей. Каждый пиксель p на экране представляется парой координат $p(x, y)$ и может иметь только положительные значения.

1.3.2. Система координат объекта

Система координат объекта Q используется в трехмерной графике для описания его геометрической модели. Поэтому данную систему иногда называют объектной или модельной. Так как в данном случае рассматриваются трехмерные объекты, то, естественно, система координат имеет три измерения. Расположение начала системы координат и направление ее осей определяются конкретно для каждого объекта исходя из требований задачи.

1.3.3. Система координат камеры

Эта система координат C привязана к камере, и она определяет расположение камеры относительно сцены. В зависимости от ориентации камеры изображение сцены имеет разное представление. В связи с этим важным является выбор такого положения камеры в каждом случае, чтобы оно максимально способствовало решению поставленной задачи.

1.3.4. Декартова система координат изображения

Данная система координат D отличается от экранной системы расположением ее начала, которое помещается в середине экрана. Направление осей – как и у обычной декартовой системы координат. Поэтому значения координат пикселей могут принимать как положительные, так и отрицательные величины. Еще одно отличие заключается в том, что на эти координаты не накладывается ограничение их целочисленности. Пиксель может представляться трехкомпонентным вектором, в котором третья компонента имеет значение интенсивности пикселя.

1.3.5. Мировая система координат

Мировая система координат W нужна для описания объектов трехмерных сцен. Многие алгоритмы трехмерной графики и компьютерного зрения работают в так называемом объектном пространстве и оперируют координатами мировой системы координат.

На рисунке 2 показаны описанные системы координат. Вершина A куба может быть определена своими координатами в каждой из них.

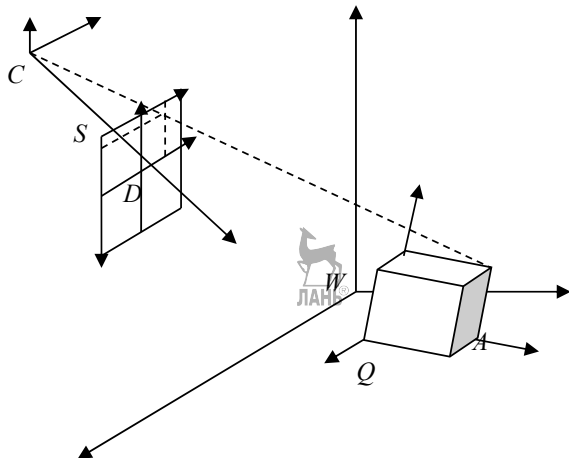


Рис. 2. Системы координат



2. Анализ бинарных изображений

В большинстве задач компьютерного зрения и трехмерной графики используются полноцветные изображения. Однако в процессе решения частных задач нередко возникает необходимость перехода к обработке полутоновых или монохромных изображений. Полутоновые изображения обычно имеют 256 степеней градации яркости, а монохромные, как правило, черно-белые (бинарные), – только два уровня, обозначаемые 0 или 1.

2.1. Обработка черно-белых изображений

Работа с двухцветными изображениями привлекает разработчиков достаточно простыми алгоритмами, и если удастся решить какую-то часть задачи с помощью бинарного изображения, то такая методика является предпочтительней по сравнению с другими подходами.

Получение бинарного изображения можно выполнить по цветному или полутоновому изображению. В зависимости от решаемой задачи может выбираться и алгоритм перевода исходного изображения в черно-белое. Наиболее распространен метод пороговой бинаризации. В этом случае определяется некоторый порог (диапазон) яркости или цвета, в соответствии с которым каждый пиксель переводится в значение 0 или 1.

При обработке или анализе изображений рассматривают не только обрабатываемый пиксель, но и соседние с ним пиксели. Здесь возможны два варианта такого соседства пикселей, определяемых связностью рассматриваемых областей. В этом смысле области могут быть 4- или 8-связными. То есть у каждого пикселя может быть 4 или 8 соседних пикселей (рис. 3).

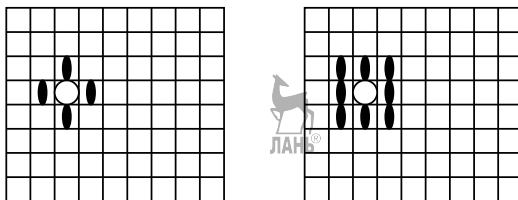


Рис. 3. Примеры 4- и 8-связных областей

Наиболее распространенный метод обработки изображений, в том числе и бинарных, связан с использованием фильтров (масок). Фильтры

представляют собой некую матрицу коэффициентов, обычно квадратную. Размер матрицы обязательно с нечетным числом элементов 3×3, 5×5 и т. д. Этой матрице соответствует окно (апертура), накладываемое на исходное изображение. Обработка заключается в умножении значений пикселей, попавших в окно, на соответствующие коэффициенты. Полученные произведения суммируются и делятся на нормирующий коэффициент, как правило, равный сумме коэффициентов фильтра. Полученный результат присваивается в качестве нового значения пикселя, обычно – центрального. Эта процедура представляет собой шаг обработки одного пикселя. После этого выполняется переход к обработке следующего пикселя. Таким образом, после обработки всех пикселей изображения получается новое результирующее изображение. Иногда полученное изображение может быть подвергнуто новой обработкой с помощью того же фильтра. Такие фильтры называются рекурсивными, т.е. полученное в результате обработки изображение становится исходным для следующей обработки.

Широкий набор фильтров позволяет получать различные изображения, способствующие решению различных задач распознавания и компьютерного зрения.

2.2. Подсчет объектов на изображении

Исходным является бинарное изображение, содержащее некоторые объекты, цвет пикселей которых обозначается значением 0, а фон – цветом 1. Объекты представляются сплошными областями, не содержащими дыр. Каждый прямоугольный объект имеет 4 внешних угла. Если все объекты имеют форму прямоугольника, то достаточно подсчитать число внешних углов и разделить это число на 4. Таков алгоритм подсчета числа объектов на изображении для простейшего случая. Если объект имеет более сложную конфигурацию, то кроме внешних углов появляются и внутренние углы. При этом число внешних углов увеличится на величину, равную числу внутренних углов.

Поэтому алгоритм подсчета объектов для общего случая [1] заключается в подсчете числа внешних углов E и числа внутренних I . Количество объектов определится как

$$Q = (E - I)/4.$$

Каждый пиксель совмещается с левым верхним элементом маски.

Подсчет внешних углов можно выполнить за счет использования масок следующего вида (рис. 4).

0	0
0	1

0	0
1	0

1	0
0	0

0	1
0	0

Рис. 4. Маски для подсчета внешних углов

Необходимо проверить на совпадение этого пикселя и трех его соседей с одной из четырех указанных на рисунке масок. Если произойдет совпадение с какой-то маской, то данный пиксель представляет собой внешний угол объекта. Аналогично подсчитывается число внутренних углов объектов по следующим маскам (рис. 5):

1	1
1	0

1	1
0	1

0	1
1	1

1	0
1	1

Рис. 5. Маски для подсчета внутренних углов

Ниже приводится алгоритм подсчета объектов на бинарном изображении на псевдокоде.

```

{
E = 0;
I = 0;
for L = 0 to MaxX - 1
  for P = 0 to MaxY - 1
    {
      if ext_match(L, P) then E = E + 1;
      if int_match(L, P) then I = I + 1;
    }
Q = (E - I)/4;
}

```

Здесь требуется детализация простых операций определения внешних и внутренних углов, которые достигаются организацией двойного цикла сравнения четырех пикселей каждой из четырех масок и формированием значения true, если четыре пикселя совпадут со значениями одной из масок, и false – в противном случае.

2.3. Маркировка связных компонент изображения

Алгоритм маркировки связных компонент бинарного изображения заключается в выделении всех связных компонент и присвоением каждой из них своей метки (идентификатора). Эти метки отличают компоненты между собой. Связными компонентами являются такие области изображения, в которых можно найти путь между двумя любыми пикселями, то есть они имеют связь между собой. Нахождение пути выполняется с перемещениями по четырем направлениям для 4-связных компонент и по восьми – для 8-связных компонент.

Известны разные алгоритмы решения указанной задачи. Рассмотрим в качестве примера рекурсивный алгоритм [1]. Дается бинарное изображение В размером $\text{MaxX}+1$ и $\text{MaxY}+1$. Необходимо получить маркированное изображение LB, в котором каждому пикселю присвоена метка компоненты, к которой он принадлежит.

Алгоритм строится следующим образом. На первом этапе выполняют смену знака у всех пикселей со значением 1. Это необходимо для различения необработанных пикселей (значение -1) и пикселей, помеченных идентификаторами. При этом изображение В переводится в текущее значение изображения LB. Затем выполняется поиск пикселя со значением -1 . Ему присваивается значение очередной метки. Для каждого соседнего пикселя рекурсивно повторяется операция присвоения метки. Алгоритм маркирования бинарного изображения, записанный на псевдокоде, приведен ниже.

```
procedure recursive_connected_components(B, LB);  
{  
  LB = negate(B);  
  label = 0;  
  find_components(LB, label);  
  print (LB);  
}
```

```
procedure find_components(LB, label);  
{  
  for L = 0 to MaxX  
    for P = 0 MaxY  
      if LB[L, P] == -1 then  
        {
```

```

        label = label + 1;
        search(LB, label, L, P);
    }
}

```

```

procedure search((LB, label, L, P);
{
  LB[L, P] = label;
  Nset = neighbors(L,P);
  for each [L', P'] in Nset
    {
      if LB[L', P'] == -1
        then search(LB, label, L', P');
    }
}

```



В приведенной процедуре следует только уточнить несложные процедуры присвоения отрицательных значений negate, выбора соседних пикселей neighbors и печати print.

2.4. Морфология бинарных изображений

Операция определения морфологии бинарных изображений имеет своей целью определения свойств формы и структуры каких-либо объектов. В компьютерном зрении под морфологией бинарных изображений обычно понимают описание свойств формы областей на изображении. Применяется операция, как правило, после маркировки бинарных изображений. Для выполнения морфологии изображений используют некоторый набор операций и структурирующие элементы.

2.4.1. Структурирующие элементы

Для выполнения операций бинарной морфологии необходимо бинарное изображение B и структурирующий элемент S (structuring element). Структурирующий элемент представляет собой некоторую область в виде бинарного изображения, которая описывает форму какого-либо объекта. Существует набор наиболее часто используемых стандартных структурирующих элементов, дополняемый при необходимости другими элементами в зависимости от решаемой задачи. Примерами структурирующих элементов могут быть области в виде прямоугольника, эллипса, тора, уголка, двутавра, полосы и т. д.

Один из пикселей структурирующего элемента выбирается в качестве начала координат, он является точкой привязки при наложении его на обрабатываемое изображение. Такой точкой можно выбирать центр элемента, угловую точку или вообще любой пиксель. Используя начало координат, можно накладывать структурирующий элемент на исходное изображение для выполнения какой-либо морфологической операции с последующим анализом результата.

2.4.2. Основные операции бинарной морфологии

К основным операциям относят операции наращивания (dilation), эрозии (erosion), замыкания (closing) и размыкания (opening). Смысл этих бинарных операций заключается в увеличении области изображения операцией наращивания, уменьшения – операцией эрозии, заливки дыр внутри областей – операцией замыкания и удаления несущественных выступающих деталей на изображении сплошных областей. Поясним более подробно суть выполнения операций.

Операция **наращивания** математически определяется как операция объединения:

$$B \oplus S = \bigcup_{b \in B} S_b .$$

Она применяется для всех пикселей исходного бинарного изображения B . Для этого начало координат структурирующего элемента совмещается с очередным пикселем изображения B . Затем выполняется операция объединения, и результат записывается в выходное изображение. После этого производится обработка следующего пикселя. По завершении обработки всех пикселей получают итоговое бинарное изображение.

Операция **эрозии** бинарного изображения B структурирующим элементом S определяется как

$$B - S = \{b \mid b + s \in B \forall s \in S\} .$$

При выполнении этой операции также выполняется обработка каждого пикселя изображения B наложением структурирующего элемента. Если в некотором положении каждый единичный пиксель структурирующего элемента совпадает с единичным пикселем бинарного изображения, то выполняется логическое сложение центрального пикселя структурирующего элемента с соответствующим пикселем выходного изображения.

Операция **замыкания** бинарного изображения задается выражением

$$B \bullet S = (B \oplus S) - S .$$

Операция **размыкания** бинарного изображения задается выражением:

$$B \circ S = (B - S) \oplus S.$$

Кроме описанных морфологических операций возможно использование и других бинарных операций в зависимости от требований практической задачи. Бинарная морфология может использоваться для извлечения характерных признаков объекта, которые затем могут служить информацией для распознавания объектов. Такими признаками, которые получаются с помощью бинарной морфологии, могут быть конфигурация объектов, геометрические размеры, пропорции его частей, периметр, центр тяжести и другие.

Еще одним примером использования бинарной морфологии является выделение на бинарном изображении компонент, у которых форма и размеры удовлетворяют заданным параметрам. В этих случаях можно подобрать соответствующую форму структурирующего элемента, с помощью которого можно получить необходимый результат. При этом для дальнейшего анализа могут потребоваться и необработанные компоненты. Для таких случаев была введена еще одна операция под названием условного наращивания (conditional dilation). Определим операцию условного наращивания. Пусть B – исходное бинарное изображение, C – обработанное бинарное изображение, S – структурирующий элемент. И пусть $C_0 = C_n$, и

$$C_n = (C_{n-1} \oplus S) \cap B.$$

Условным наращиванием изображения C элементом S по отношению к B называется операция

$$C \oplus_{|B|} S = C_m,$$

где индекс m равен наименьшему значению, при котором $C_m = C_{m-1}$.

Согласно определению множество $C = C_0$ циклически наращивается структурирующим элементом S , и на каждом шаге результат уменьшается до подмножества пикселей, которые имеют единичные значения на исходном изображении B .

2.5. Свойства бинарных областей

После обработки бинарного изображения и получения множества бинарных областей нередко требуется проведение анализа этих областей с целью получения свойств областей. Полученные результаты анализа будут служить исходной информацией для последующего решения задач

компьютерного зрения. Примерами таких наиболее распространенных свойств могут быть:

- геометрические свойства – площадь области, размеры, центр тяжести;
- свойства форм областей – прямоугольные, округлые, вытянутые, зигзагообразные;
- свойства интенсивности – яркие, темные, средние;
- свойства текстуры – характерные рисунки, повторяющиеся узоры, статистика областей.

Ниже приводятся определения некоторых часто используемых геометрических свойств и свойств форм областей. В этих определениях множество пикселей области обозначается буквой R , и каждый пиксель считается квадратным.

2.5.1. Площадь области и центр тяжести

Площадь области A определяется суммой единичных пикселей области

$$A = \sum_{(x,y) \in R} 1.$$

Определение координат $(x_{\text{цт}}, y_{\text{цт}})$ центра тяжести области вычисляется следующим образом:

$$x_{\text{цт}} = \frac{1}{A} \sum_{(x,y) \in R} x;$$
$$y_{\text{цт}} = \frac{1}{A} \sum_{(x,y) \in R} y.$$

2.5.2. Длина периметра

Периметр области представляет собой множество граничных пикселей области. В этом определении предполагается отсутствие внутренних отверстий в областях. Пиксель области R является граничным, если один из его соседей не принадлежит R . Для четырехсвязных областей граничные пиксели образуют восьмисвязные области, и, наоборот, у восьмисвязных областей граничные пиксели четырехсвязные. Множества пикселей периметра P_8 для 4- и P_4 для 8-связных областей определяются следующим образом:

$$P_8 = \{(x, y) \in R \mid M_4(x, y) - R \neq 0\};$$

$$P_4 = \{(x, y) \in R \mid N_8(x, y) - R \neq 0\}.$$

Длина периметра $|P|$ вычисляется по упорядоченной цепочке соседних пикселей периметра $P = \langle (x_0, y_0), (x_1, y_1), \dots, (x_{k-1}, y_{k-1}) \rangle$ по следующему выражению:

$$|P| = |\{k \mid (x_{k+1}, y_{k+1}) \in N_4(x_k, y_k)\}| + \sqrt{2} |\{k \mid (x_{k+1}, y_{k+1}) \in N_8(x_k, y_k) - N_4(x_k, y_k)\}|,$$

где k – длина последовательности пикселей;

$k + 1$ – вычисляется по модулю K (K равно длине последовательности пикселей).

В соответствии с этим определением длины периметра два соседних четырехсвязных пикселя увеличивают длину периметра на единицу, а восьмисвязных – на величину $\sqrt{2}$.

2.5.3. Свойство округлости

На основании понятия площади A и периметра P вводится понятие округлости (circularity) области как квадрата длины периметра, деленного на площадь области:

$$C_1 = \frac{|P|^2}{A}.$$

Ввиду того, что значение округлости имеет различные значения для одной и той же области в зависимости от определения ее 4- или 8-связной, а также от того, что она принимает минимальное значение не для окружности, а для ромба и восьмиугольников, было предложено [5] другое определение округлости:

$$C_2 = \frac{\mu_R}{\sigma_R},$$

где μ_R – среднее отклонение граничных пикселей от центра тяжести области;

σ_R – среднеквадратичное отклонение граничных пикселей от центра тяжести области.

Указанные значения отклонений рассчитываются по формулам, приведенным ниже. Здесь множество пикселей (x_k, y_k) , $k = 0, \dots, K-1$ принадлежит периметру P рассматриваемой области. Второе определение округлости ведет себя похожим образом для цифровых и непрерывных образов фигур:

$$\mu_R = \frac{1}{K} \sum_0^{K-1} \|(x_k, y_k) - (x_{\text{цт}}, y_{\text{цт}})\|;$$

$$\sigma_R = \left(\frac{1}{K} \sum_0^{K-1} [\|(x_k, y_k) - (x, y)\| - \mu_R]^2 \right)^{1/2}.$$

2.5.4. Описывающий прямоугольник и экстремальные точки

Для каждой области бинарного изображения можно построить описывающий прямоугольник в виде минимаксной оболочки по минимальным и максимальным значениям координат x и y . Эта оболочка для i области будет иметь следующие координаты:

$$x_{\min}(i), x_{\max}(i), y_{\min}(i), y_{\max}(i).$$

Прямоугольная оболочка будет иметь стороны, параллельные осям координат.

На каждой стороне описывающего прямоугольника в общем случае можно отметить по две экстремальные точки: на верхней и нижней стороне – крайние левые и крайние правые, на левой и правой стороне – крайние верхние и крайние нижние. Эти экстремальные точки обозначают контур бинарной области (рис. 6).

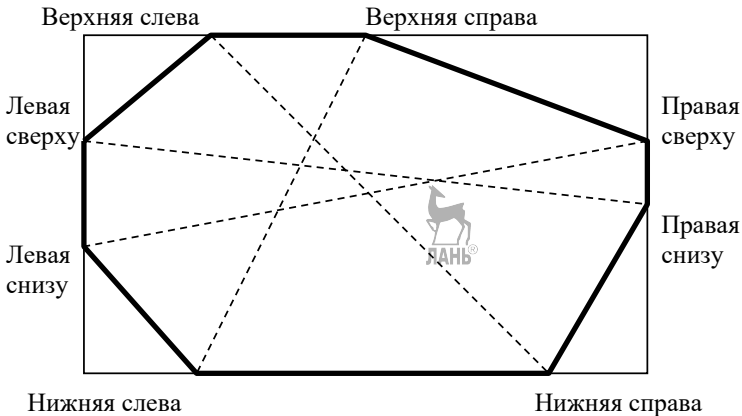


Рис. 6. Экстремальные точки описывающего многоугольника

Линии соединения противоположных экстремальных точек дают соединительные оси, для которых, в некоторых случаях, может быть полезно определить длину и направление.

Длина экстремальной оси может быть вычислена так:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + Q(\theta),$$

где $Q(\theta)$ – поправка, учитывающая погрешность перехода к целочисленным координатам;

θ – угол, определяющий направление экстремальной оси.

Значения поправки лежат в интервале от 0 до 1 и определяются как

$$Q(\theta) = \begin{cases} \frac{1}{|\cos \theta|}; & |\theta| < 45^\circ; \\ \frac{1}{|\sin \theta|}; & |\theta| > 45^\circ. \end{cases}$$

2.5.5. Моменты второго порядка для областей

Для описания формы области применяют и моменты второго порядка. Такими моментами являются центральные моменты по строке, столбцу и смешанные. Вычисляются эти моменты по нижеприведенным формулам.

Центральный момент второго порядка по строке:

$$\mu_{xx} = \frac{1}{A} \sum_{(x,y) \in R} (x - x_{\text{цт}})^2.$$

Центральный момент второго порядка по столбцу:

$$\mu_{yy} = \frac{1}{A} \sum_{(x,y) \in R} (y - y_{\text{цт}})^2.$$

Смешанный центральный момент второго порядка:

$$\mu_{xy} = \frac{1}{A} \sum_{(x,y) \in R} (x - x_{\text{цт}})(y - y_{\text{цт}}).$$

У некоторых битовых областей центральные моменты относительно некоторых направлений могут сильно различаться. В этом случае можно выделить главную ось, которая будет являться характерной осью по сравнению с другими. Главной осью объекта на плоском бинарном изображении будет та ось, относительно которой пиксели области имеют наименьший момент второго порядка. По аналогии с механикой эта ось является осью, относительно которой момент инерции объекта минимален. Можно заметить, что момент инерции кругового диска относительно всех осей имеет одинаковое значение.

Известно, что главная ось, соответствующая минимальному моменту инерции, проходит через центр тяжести $(x_{\text{цт}}, y_{\text{цт}})$ множества пикселей области. Множество моментов относительно заданного набора осей может служить набором дополнительных характеристик бинарных изображений при распознавании образов. Для этого используется формула расчета моментов второго порядка для произвольной оси.

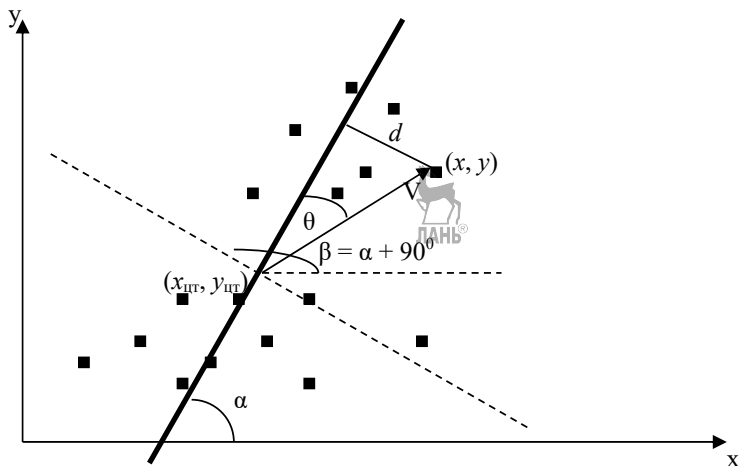


Рис. 7. Пример бинарной области с осью, проходящей через центр тяжести

На рисунке 7 представлена бинарная область с указанием произвольной оси, направленной через центр тяжести под углом α . Для определения момента второго порядка множества пикселей, относительно указанной оси, необходимо для всех пикселей вычислить сумму квадратов расстояний d до оси и выполнить нормировку делением на количество пикселей области. Здесь d определяется как проекция вектора V на ось, перпендикулярную к выбранной оси ($\beta = \alpha + 90^\circ$), через скалярное произведение векторов:

$$\begin{aligned} \mu(x_{\text{цт}}, y_{\text{цт}}, \alpha) &= \frac{1}{A} \sum_{(x,y) \in R} d^2 = \frac{1}{A} \sum_{(x,y) \in R} (V \circ (\cos \beta, \sin \beta))^2 = \\ &= \frac{1}{A} \sum_{(x,y) \in R} ((x - x_{\text{цт}}) \cos \beta + (y - y_{\text{цт}}) \sin \beta)^2. \end{aligned}$$

С помощью приведенной формулы можно считать моменты второго порядка относительно любых осей. Часто используются такие моменты относительно горизонтальных, вертикальных или диагональных осей, например, для распознавания текста.

Наибольший интерес представляют моменты второго порядка относительно осей с углами α^* , дающих экстремальные значения момента

инерции. Для этого необходимо исследовать выражение для $\mu(x_{\text{ит}}, y_{\text{ит}}, \alpha)$ на экстремум путем дифференцирования по углу α :

$$\begin{aligned} \operatorname{tg} 2\alpha^* &= \frac{2\sum(x-x_{\text{ит}})(y-y_{\text{ит}})}{\sum(x-x_{\text{ит}})(x-x_{\text{ит}}) - \sum(y-y_{\text{ит}})(y-y_{\text{ит}})} = \\ &= \frac{\frac{1}{A}2\sum(x-x_{\text{ит}})(y-y_{\text{ит}})}{\frac{1}{A}\sum(x-x_{\text{ит}})(x-x_{\text{ит}}) - \frac{1}{A}\sum(y-y_{\text{ит}})(y-y_{\text{ит}})} = \\ &= \frac{2\mu_{xy}}{\mu_{xx} - \mu_{yy}}. \end{aligned}$$

Углы осей α^* с максимальными и минимальными значениями моментов инерции различаются на 90° . В каждом конкретном случае их можно легко различить.



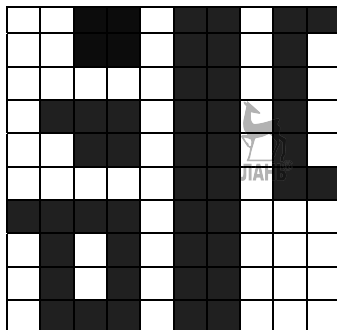
2.6. Графы смежности областей

При анализе бинарных изображений, кроме свойств областей, интерес представляет структура распределения областей. На изображении рассматриваются области двух видов: переднего плана и фона. Ранее рассматривался алгоритм маркировки областей переднего плана. Аналогично может быть поставлена задача маркировки и областей фона. Если маркировка переднего плана использовала натуральный ряд чисел, то удобно промаркировать области фона числами $0, -1, -2$ и т. д.

После выполнения маркировки можно построить граф смежности областей, в котором вершины соответствуют областям переднего плана и фона, а ребрами – соседство (смежность) областей. Области считаются смежными, если они имеют соседние пиксели. Причем соседство может рассматриваться как для 4-, так и 8-связных компонент.

Алгоритм построения графа смежности областей заключается в анализе изображения по две строки. Такой подход позволяет обнаруживать смежность областей. В случае наличия соседства двух разных областей переднего плана и фона в граф добавляется соответствующая связь для этих областей.

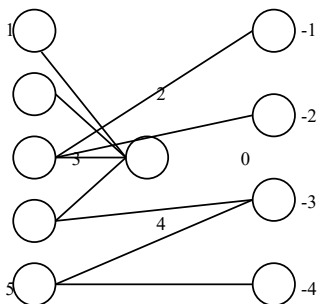
Ниже на рисунке 8 приведен пример бинарного изображения, маркировки и графа смежности областей.



Бинарное изображение

0	0	1	1	0	4	4	-3	5	5
0	0	1	1	0	4	4	-3	5	-4
0	0	0	0	0	4	4	-3	5	-4
0	2	2	2	0	4	4	-3	5	-4
0	0	2	2	0	4	4	-3	5	-4
0	0	0	0	0	4	4	-3	5	5
3	3	3	3	0	4	4	-3	-3	-3
-1	3	-2	3	0	4	4	-3	-3	-3
-1	3	-2	3	0	4	4	-3	-3	-3
-1	3	3	3	0	4	4	-3	-3	-3

Маркировка областей



Граф смежности

Рис. 8. Пример маркировки и графа смежности бинарного изображения

2.7. Пороговая бинаризация изображений

Бинарное изображение обычно получают по полутоновому изображению. Поэтому при наличии цветного изображения ведут его обработку последовательно: цветное – полутоновое – бинарное. Первое преобразование цвет – полутон выполняется вычислением среднеарифметического значения по трем базовым цветам. Преобразование полутонового изображения в бинарное изображение может выполняться по одному из трех вариантов:

- по верхнему пороговому значению;
- нижнему пороговому значению;
- диапазону.

Например, в первом случае, если значение яркости превышает порог, то данный пиксель считается пикселем переднего плана. Аналогично поступают и в остальных случаях. В этой процедуре важным является выбор порогового значения или диапазона, так как это существенно влияет на получение конечного результата решения задачи распознавания.

Выбор порога пользователь может выполнять произвольно методом проб и ошибок или используя гистограмму изображения. В последнем случае возможен выбор порога, который хорошо различает границу раздела светлых и темных пикселей. Известны методы автоматического выбора порогового значения бинаризации. В некоторых случаях выполняют динамическую пороговую бинаризацию, при которой для разных частей изображения подбирают разные пороги.

Наконец, существуют приемы пороговой бинаризации на основе априорной информации. В этом случае порог бинаризации выбирается на основе предварительной информации о распознаваемых объектах и характере их расположения на изображении.



3. Обработка полутоновых изображений

Обработка изображений предполагает получение выходного изображения в результате выполнения одной или нескольких операций над исходным изображением. В качестве операций обработки изображения могут выполняться фильтрации, улучшения изображений, изменения качественных характеристик, выделения каких-то параметров и т. д. Обработка изображения может основываться на анализе окрестности отдельного пикселя или на основе характеристик всего изображения.

Если функция преобразования изображения определяет новое значение пикселя по значению одного исходного пикселя, то она является точечным оператором. Примером такого точечного оператора является оператор усиления контраста, в результате применения которого усиливаются детали изображения. Важным свойством точечных операторов является возможность независимой обработки отдельных пикселей. Это позволяет выполнять параллельную обработку изображения.

В данной главе рассматриваются различные методы улучшения изображения операциями подавления шумов, сглаживания, повышения контраста и усиления краев.

3.1. Увеличение диапазона полутонов

Увеличение диапазона интенсивности изображения существенно улучшает качество изображения за счет усиления контраста. Для этого используют три метода – линейная растяжка, нормализация и выравнивание гистограмм. Все эти методы направлены на расширение диапазона интенсивности до максимального интервала от 0 до 255.

Линейная растяжка гистограммы заключается в преобразовании уровней яркости исходного изображения. Преобразование выполняется по формуле

$$g_i = a + b * f_i,$$

где f_i – значение яркости i -го пикселя исходного изображения;

g_i – значение яркости выходного изображения;

a, b – коэффициенты линейной растяжки.

При таком подходе диапазон яркости исходного изображения от f_{\min} до f_{\max} преобразуется в диапазон от 0 до 255. Исходя из этого коэффициенты a, b могут быть рассчитаны по формулам:

$$a = 255 \frac{f_{\min}}{f_{\min} - f_{\max}}; \quad b = \frac{255}{f_{\max} - f_{\min}}.$$

Нормализация гистограммы также предусматривает растяжку гистограммы, но не всего диапазона яркости исходного изображения (от f_{\min} до f_{\max}), а только ее части – наиболее интенсивного участка (от f_{\min} до f_{\max}). При этом исключают из рассмотрения малоинформативные «хвосты». Поэтому предварительно необходимо установить преобразуемый интервал яркости ($f_{\min} - f_{\max}$), а затем выполнить преобразование по тем же формулам линейной растяжки.

Выравнивание гистограммы (эквализация) ставит своей задачей такое преобразование гистограммы, после которого все уровни яркости имеют примерно одинаковую частоту. При этом гистограмма яркостей будет иметь равномерный закон распределения.

Эту операцию выполняют с условием использования всех допустимых значений интенсивности (использование всего диапазона), и выходное изображение должно иметь примерно равное количество пикселей для каждого значения интенсивности. Таким образом, в идеале гистограмма, преобразованная с указанными условиями, дает гистограмму, у которой 256 уровней, а количество пикселей каждого уровня равно $(\text{MaxX} * \text{MaxY}) / 256$.

Для определения функции преобразования f , растягивающей динамический диапазон изображения, необходима гистограмма исходного изображения. Для каждого уровня интенсивности гистограммы $H_{in}[i]$ указывается количество пикселей, имеющих значение интенсивности z_i .

Первое пороговое значение интенсивности t_1 определяется наращиванием индекса i на гистограмме входного изображения до тех пор, пока суммарное количество пикселей с меньшими значениями интенсивности не превысит порог q_1 . Всем пикселям входного изображения со значением интенсивности $z_k < t_1 - 1$ на выходном изображении будут присвоены значения z_1 . Пороговое значение t_1 будет определяться с помощью выражения

$$\sum_{i=1}^{t_1-1} H_{in}[i] \leq q_1 < \sum_{i=1}^{t_1} H_{in}[i].$$

Из выражения следует, что для интенсивности t_1 имеется не более q_1 пикселей с меньшими значениями.

В общем случае для k -го порогового значения выражение имеет вид

$$\sum_{i=1}^{t_k-1} H_{in}[i] \leq (q_1 + q_2 + \dots + q_k) < \sum_{i=1}^{t_k} H_{in}[i].$$

Для записи преобразования f обычно используют таблицу преобразования (lookup table, LUT), которая формируется на основе приведенного выше выражения.

3.2. Удаление мелких деталей

За редким исключением, мелкие детали на изображении являются шумом. Удаление таких областей повышает качество изображения. Для этого используются разные методы, которые позволяют удалять как отдельные пиксели, так и области, интерпретируемые в качестве шума на изображении. Ниже приводятся два таких алгоритма.

3.2.1. Удаление шума «соль и перец»

Шум под названием «соль и перец» представляет собой одиночные темные пиксели внутри светлых областей (перец) или одиночные светлые пиксели внутри темных областей (соль). Он может возникать в результате пороговой бинаризации изображения, а также после аналого-цифровых преобразований. Возможны и другие случаи возникновения шума. Удаление шума «соль и перец», как правило, основано на применении масок, с помощью которых выявляется наличие шума, и масок, которые удаляют этот шум. На рисунках 9 и 10 показаны маски для случаев бинарного изображения и бинарного маркированного изображения. Кроме того, указаны маски для 4- или 8-связных областей.



Рис. 9. Выявление и устранение шума на бинарном изображении



Рис. 10. Выявление и устранение шума на бинарном маркированном изображении 4- и 8-связных областей

Здесь левые маски в каждой паре служат для выявления того или иного случая, а правые – для удаления шума.

3.2.2. Удаление мелких областей

Удаление мелких областей, которые могут быть отнесены к шумам, выполняется на основе расчета их признаков. Такими признаками могут быть размер области (площадь, количество пикселей), геометрические

особенности (вытянутость, маленький размер одного из измерений габаритов). В каждом конкретном случае требуется установить набор таких параметров. После выделения компонент связности битового изображения выполняется удаление тех областей, которые по своим свойствам подпадают под определение шума. В некоторых случаях может требоваться сохранение информации об удаленных областях, если возникнет необходимость возврата к исходному изображению.

3.2.3. Сглаживание изображений

Если изображение содержит кроме основного идеального изображения некоторый случайный шум типа гауссова шума, то улучшение изображения можно достичь применением сглаживающего фильтра. Гауссовый шум описывается нормальным распределением с математическим ожиданием, равным нулю, и среднеквадратичным отклонением σ . Самый простой вариант сглаживающего фильтра содержит единицы для каждого пикселя попадающего в окно апертуры. Размер окна обычно может быть 3×3 или 5×5 . При наложении окна на изображение выполняется расчет нового значения пикселя, который находится в центре окна. Для этого выполняется суммирование всех пикселей в окне и деление полученной суммы на 9 (окно 3×3) или 25 (окно 5×5). То есть новое значение представляет собой среднее арифметическое значение всех пикселей в окне.

Лучшее качество сглаживания получается при использовании фильтра Гаусса, который отличается от описанного тем, что весовые коэффициенты фильтра зависят от расстояния до центрального пикселя. Эти коэффициенты уменьшаются с увеличением расстояния. В этом случае наибольшее влияние на результат оказывают ближние пиксели, а наименьшее – удаленные. Весовая функция рассчитывается по формуле

$$g(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d^2}{2\sigma^2}},$$

где $d = \sqrt{(x - x_c)^2 + (y - y_c)^2}$ – расстояние от пикселя (x, y) до центрального пикселя (x_c, y_c) .

3.2.4. Медианный фильтр

Устранение шумов сглаживающим фильтром в некоторых случаях приводит к нежелательным результатам, например к размытию границ или контуров. Поэтому в таких случаях широко используется медианный фильтр, который заменяет значение обрабатываемого пикселя на медиану

ну значений пикселей из его окрестности. Медиана представляет собой значение срединного элемента упорядоченного массива. Если задан упорядоченный массив значений

$$A = (A_0, A_1, \dots, A_n),$$

то медианой этого массива будет значение элемента с индексом $n/2$. При четном числе элементов массива медианой может считаться один из элементов с номерами $(n - 1)/2$ или $(n + 1)/2$. Иногда медианой считают среднее значение из этих двух элементов.

Применение медианного фильтра позволяет устранять шумы и сохранять структуру границ на изображении. Особенность этого метода сглаживания состоит в достаточно большом объеме вычислений за счет сортировки массивов.



3.3. Обнаружение краев с помощью дифференциальных масок

Границы раздела областей (края) на изображении обычно определяются вычислением разности значений интенсивности в локальных областях изображения. Такие области разграничения можно выявлять с помощью масок.

3.3.1. Дифференцирование одномерных сигналов

Пусть задан одномерный сигнал в виде числовой последовательности значений некоторой функции f . Тогда производная для такого одномерного сигнала определяется как

$$f'(x_i) \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}.$$

Если интервал отчета значений функции принять $\Delta x = 1$, тогда производная определяется разностью двух последовательных отчетов. В этом случае подсчет производной одномерного сигнала можно выполнять с помощью маски $M' = [-1, +1]$. Если в последовательности отчетов функции f взять два соседних отчета и умножить их значения на маску M' и сложить, то получится значение производной на этом интервале

$$f'(x_i) \approx -f(x_{i-1}) + f(x_i).$$

Аналогично можно рассчитать и вторую производную, используя маску M'' . Для второй производной сигнала имеем

$$\begin{aligned} f''(x_i) &= f'(x_i) - f'(x_{i-1}) = f(x_i) - f(x_{i-1}) - f(x_{i-1}) + f(x_{i-2}) = \\ &= f(x_i) - 2f(x_{i-1}) + f(x_{i-2}). \end{aligned}$$

В этом случае маска будет иметь вид $M'' = [+1, -2, +1]$. Следует отметить, что при вычислении абсолютных значений производных можно использовать и маски с инвертированными значениями коэффициентов:

$$M' = [+1, -1]; M'' = [-1, +2, -1].$$

Такие абсолютные значения производных используются для выявления точек контраста без определения направления смены знака производной.

Кроме указанных масок иногда используется и маска для вычисления первой производной из трех элементов $M = [-1, 0, +1]$, которая используется для последовательных отчетов с номерами $(i - 1), i, (i + 1)$. Так как $\Delta x = 2$, то вычисленный результат с применением этой маски следует разделить на 2.

Ниже рассматриваются примеры использования масок для выявления различных случаев границы областей.

1. Ступенчатый край, направленный вверх.

S_i	12	12	12	12	12	24	24	24	24	24
$M \times S_i$	0	0	0	0	12	12	0	0	0	0

2. Ступенчатый край, направленный вниз.

S_i	24	24	24	24	24	12	12	12	12	12
$M \times S_i$	0	0	0	0	-12	-12	0	0	0	0

3. Наклонный край, направленный вверх.

S_i	12	12	12	12	15	18	21	24	24	24
$M \times S_i$	0	0	0	3	6	6	6	3	0	0

4. Яркостный импульс.

S_i	12	12	12	12	24	12	12	12	12	12
$M \times S_i$	0	0	0	12	0	-12	0	0	0	0

Из этих примеров видно, что маска первой производной из трех элементов дает отклик в двух дискретных элементах для случая ступенчатого края.

Рассмотрим отклик второй производной для тех же случаев. Маска $M = [-1, +2, -1]$.

1. Ступенчатый край, направленный вверх.

S_i	12	12	12	12	12	24	24	24	24	24
$M \times S_i$	0	0	0	0	-12	12	0	0	0	0

2. Ступенчатый край, направленный вниз.

S_i	24	24	24	24	24	12	12	12	12	12
$M \times S_i$	0	0	0	0	12	-12	0	0	0	0

3. Наклонный край, направленный вверх.

S_i	12	12	12	12	15	18	21	24	24	24
$M \times S_i$	0	0	0	-3	0	0	0	3	0	0

4. Яркостный импульс.

S_i	12	12	12	12	24	12	12	12	12	12
$M \times S_i$	0	0	0	-12	24	-12	0	0	0	0

Для второй производной характерным является то, что для контраста отклик дает значения с переходом через ноль. Совместное использование первой и второй производных позволяет выявить ряд локальных особенностей исходного сигнала.

Аналогичное применение масок используется для сглаживания сигнала. Маски имеют трехкомпонентный размер, и сумма ее элементов равна 1. Например, усредняющая маска $M = [1/3, 1/3, 1/3]$. Ниже приведен результат ее применения для ступенчатого края и яркостного импульса.

1. Ступенчатый край, направленный вверх.

S_i	12	12	12	12	12	24	24	24	24	24
$M \times S_i$	12	12	12	12	16	20	24	24	24	24

2. Яркостный импульс.

S_i	12	12	12	12	24	12	12	12	12	12
$M \times S_i$	12	12	12	16	16	16	12	12	12	12

Маска гауссова сглаживания $M = [1/4, 1/2, 1/4]$ для тех же случаев дает следующий результат:

1. Ступенчатый край, направленный вверх.

S_i	12	12	12	12	12	24	24	24	24	24
$M \times S_i$	12	12	12	12	15	21	24	24	24	24

2. Яркостный импульс.

S_i	12	12	12	12	24	12	12	12	12	12
$M \times S_i$	12	12	12	15	18	15	12	12	12	12

Можно сформулировать некоторые свойства масок для вычисления производных:

- производные в областях сигнала с высоким контрастом имеют разные знаки;

- сумма элементов масок производных равна 0, поэтому в областях постоянного сигнала они дают нулевой отклик;
- точкам высокого контраста соответствуют большие абсолютные значения отклика масок первой производной;
- отклик масок второй производной для высокого контраста дает значения, переходящие через ноль.

Свойства сглаживающих масок:

- элементы сглаживающих масок положительны и в сумме дают единицу; поэтому отклик масок на постоянные значения сигнала совпадает с исходными значениями;
- степень сглаживания и снижения шума пропорциональна размеру маски;
- ступенчатые края сглаживаются пропорционально размеру маски.

3.3.2. Дифференцирование двумерных изображений

Направление контраста функции интенсивности f двумерного изображения определяется вычислением градиента $\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]$. Для дискретной

функции частные производные заменяются разностями изменения значений в окрестности пикселя. С целью устранения шума и локальных всплесков используют окрестности размером 3×3 . В связи с этим градиент по двум направлениям будет вычисляться по следующим формулам:

$$\begin{aligned} \frac{\partial f}{\partial x} = f_x \approx & \frac{1}{3} [(I(x+1, y) - I(x-1, y)) / 2 + \\ & + (I(x+1, y-1) - I(x-1, y-1)) / 2 + \\ & + (I(x+1, y+1) - I(x-1, y+1)) / 2]. \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial y} = f_y \approx & \frac{1}{3} [(I(x, y+1) - I(x, y-1)) / 2 + \\ & + (I(x-1, y+1) - I(x-1, y-1)) / 2 + \\ & + (I(x+1, y+1) - I(x+1, y-1)) / 2]. \end{aligned}$$

С целью упрощения вычислений можно опускать деление на 6. При этом отклик масштабируется, что не влияет на определение направления контраста.

Вычисления по указанным выражениям можно выполнять с помощью масок:

$$M_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad M_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}.$$

Иногда контраст определяется с помощью фильтра Собеля и соответствующих ему масок:

$$M_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad M_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

Здесь отличие состоит в увеличении веса центральной строки и центрального столбца.

Направление градиента и его величину с использованием приведенных выше масок определяют следующим образом:

$$\frac{\partial f}{\partial x} = f_x \approx \frac{1}{6}(M_x \circ N(x, y));$$

$$\frac{\partial f}{\partial y} = f_y \approx \frac{1}{6}(M_y \circ N(x, y));$$

$$|\nabla f| \approx \sqrt{f_x^2 + f_y^2};$$

$$\theta \approx \arctg\left(\frac{f_y}{f_x}\right).$$

Операция скалярного произведения выполняется перемножением элементов масок на соответствующие значения двумерного изображения и суммированием этих произведений.

В случаях, когда предъявляются более жесткие требования к объему вычислений, вычисления градиента можно выполнять по упрощенным формулам с небольшой потерей качества вычислений:

$$\max(|f_x|, |f_y|); \quad |f_x| + |f_y|; \quad \frac{f_x^2 + f_y^2}{2}.$$

Можно отметить, что градиентные операторы дают хорошие результаты при обнаружении границ изолированных объектов. Но эти границы могут теряться из-за их кривизны или за счет плавных переходов полутонов.

3.3.3. Обнаружение краев с помощью фильтрации

Здесь рассматривается фильтрация, использующая функцию Гаусса, и ЛОГ-фильтрация, основанная на применении оператора Лапласа к функции Гаусса.

Функция Гаусса от одной переменной имеет вид

$$g(x) = ce^{-\frac{x^2}{2\sigma^2}},$$

где σ – среднеквадратичное отклонение,

c – некоторый масштабный множитель.

Эта же функция Гаусса для двух переменных имеет следующий вид:

$$g(x, y) = ce^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Значения масштабного множителя подбираются таким образом, чтобы значения всех элементов маски для фильтрации были бы целочисленными.

Обнаружение краев выполняется путем обработки изображения поэлементно для каждого пикселя и его окрестности с помощью функции Гаусса и его первой и второй производных. Интеграл этой функции (площадь криволинейной трапеции, ограниченной функцией) равен 1. Поэтому использование функции в качестве сглаживающего фильтра не изменяет содержимого областей постоянных значений.

Если принять $c = 1/\sqrt{2\pi}$, функция Гаусса и ее производные будут иметь вид:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

$$g'(x) = \frac{-1}{\sqrt{2\pi}\sigma^3} x e^{-\frac{x^2}{2\sigma^2}} = \frac{-x}{\sigma^2} g(x).$$


$$g''(x) = \left(\frac{x^2}{\sqrt{2\pi}\sigma^5} - \frac{1}{\sqrt{2\pi}\sigma^3} \right) e^{-\frac{x^2}{2\sigma^2}} = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2} \right) g(x).$$

Используя соотношение $r = \sqrt{x^2 + y^2}$, можно построить выражение для двумерной функции $g(x, y)$ и ее производных. Эти двумерные функции имеют форму поверхностей вращения их одномерных аналогов вокруг вертикальной оси. Каждое сечение этих поверхностей будет соответствовать одномерному гауссиану.

В качестве примеров можно привести две маски аппроксимации гауссиана размерами 3×3 и 7×7 . Первая маска получена матричным перемножением вектора-строки $[1 \ 2 \ 1]$ на вектор-столбец $[1 \ 2 \ 1]$:

$$\begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array} G_{3 \times 3} =$$

Значения для второй маски вычислены по выражению



$$g(x, y) = ce^{-\frac{x^2+y^2}{2\sigma^2}}.$$


Затем выбран коэффициент $c = 90$ с тем расчетом, чтобы минимальный элемент маски был равен 1.

$$\begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 3 & 7 & 9 & 7 & 3 & 1 \\ \hline 3 & 12 & 26 & 33 & 26 & 12 & 3 \\ \hline 7 & 26 & 55 & 70 & 55 & 26 & 7 \\ \hline 9 & 33 & 70 & 90 & 70 & 33 & 9 \\ \hline 7 & 26 & 55 & 70 & 55 & 26 & 7 \\ \hline 3 & 12 & 26 & 33 & 26 & 12 & 3 \\ \hline 1 & 3 & 7 & 9 & 7 & 3 & 1 \\ \hline \end{array} G_{7 \times 7} =$$

Эти маски можно использовать для сглаживания двумерных изображений.

3.3.4. Обнаружение краев с помощью ЛОГ-фильтра

Дифференциальный фильтр второго порядка, полученный применением оператора Лапласа к функции Гаусса, называется ЛОГ-фильтром (Laplacian of the Gaussian). Приблизительно ЛОГ-фильтр можно представить в виде разности двух гауссианов:



$$g''(x) \approx c_1 e^{-\frac{x^2}{2\sigma_1^2}} - c_2 e^{-\frac{x^2}{2\sigma_2^2}}.$$

Чтобы центральная выпуклость функции лежала в положительной полуплоскости, нужно выполнять соотношение $\sigma_1 < \sigma_2$, а значение σ_2 выбирается с учетом значения σ_1 так, чтобы точки перехода через ноль среди значений отклика фильтра располагались корректно, а суммы отрицательных и положительных коэффициентов совпадали.

ЛОГ-фильтр дает значимый отклик для разности интенсивностей в двух случаях – для небольших пятен с центром, совпадающим с центром маски фильтра и для широких ступенчатых краев, расположенных близко

от центра маски. Маски малого размера, например 3×3, позволяют обнаруживать мелкие детали изображения вплоть до размера в один пиксель. Маски больших размеров детектируют более крупные детали. При этом в использовании этих масок имеется и такая особенность, как объем вычислений, который обладает значительными вычислениями для масок больших размеров.

Существует гипотеза Марра – Хильдрет, которая предполагает, что в основе человеческого зрения лежит ЛОГ-фильтрация, с помощью которой зрительная система человека строит некоторый эскиз видимой сцены, состоящий из двумерных примитивов (прямые, области, пятна). Для получения полноценной информации о наблюдаемой сцене ЛОГ-фильтрация выполняется с различными значениями σ , что позволяет проработать мелкие и крупные детали. Далее эта информация подвергается обработке на более высоком уровне, при которой происходит соединение результатов фильтрации. Наличие стереографического зрения позволяет получить и трехмерную картину изображения. Предложенная модель неплохо объясняет свойства зрения человека и животных. Однако последние исследования в этой области говорят о необходимости более сложной модели зрительной системы человека. Но, несмотря на это, методы обработки изображений с использованием гауссианов и ЛОГ-фильтрации переменного размера в настоящее время применяются достаточно широко.

3.4. Использование масок для анализа изображения

Решение различных задач распознавания связано с анализом сходства некоторой окрестности изображения с маской. С этой целью разрабатываются наборы масок, предназначенных для выделения признаков конкретного вида. Такой анализ изображений носит название согласованной фильтрации. Она может рассматриваться как для одномерных, так и двумерных изображений, считая первые частным случаем вторых.

3.4.1. Некоторые понятия векторов

Множество всех векторов, состоящих из n действительных чисел, образует векторное пространство. С каждым вектором $S = (s_1, s_2, \dots, s_n)$ связано понятие длины, определяемой как корень квадратный из суммы квадратов его координат:

$$d = \sqrt{s_1^2 + s_2^2 + \dots + s_n^2}.$$

В области обработки сигналов, кроме длины вектора, часто оперируют понятием энергии сигнала. Энергия сигнала Q равна квадрату длины вектора или сумме квадратов его координат:

$$Q = d^2 = s_1^2 + s_2^2 + \dots + s_n^2.$$

Два произвольных вектора S_1 и S_2 являются ортогональными, если их скалярное произведение равно 0:

$$S_1 \circ S_2 = 0.$$

Базис n -мерного векторного пространства представляет собой множество n векторов $\{w_1, w_2, \dots, w_n\}$, которые являются линейно независимыми, и покрывают векторное пространство. Покрытие векторного пространства заключается в том, что любой вектор V можно представить в виде линейной комбинации базисных векторов:

$$V = a_1 w_1 + a_2 w_2 + \dots + a_n w_n,$$

где a_i – числовые коэффициенты.

Свойство линейной независимости означает, что никакие из этих базисных векторов w_i нельзя представить в виде линейной комбинации остальных базисных векторов.

Из всех возможных представлений произвольного вектора особый интерес заслуживает представление вектора с помощью ортогонального базиса. Пусть задан базис из трех векторов $\{w_1, w_2, w_3\}$. Любой сигнал $S = \{a_1, a_2, a_3\}$ можно представить как линейную комбинацию базисных векторов $S = a_1 w_1 + a_2 w_2 + a_3 w_3$. Тогда скалярное произведение сигнала S на произвольный вектор из базиса будет равно

$$S \circ w_i = a_1 (w_1 \circ w_i) + a_2 (w_2 \circ w_i) + a_3 (w_3 \circ w_i) = a_i (w_i \circ w_i) = a_i.$$

Здесь

$$w_i \circ w_j = \begin{cases} 0, & \text{при } i \neq j, \\ 1, & \text{при } i = j. \end{cases}$$

С помощью стандартного ортогонального базиса единичной длины любой вектор $V = (v_1, v_2, v_3)$ можно представить в следующем виде:

$$V = (v_1, v_2, v_3) = v_1(1, 0, 0) + v_2(0, 1, 0) + v_3(0, 0, 1).$$

Рассмотрим пример другого множества базисных векторов $\{w_1, w_2, w_3\}$, у которого

$$w_1 = (-1, 0, 1); \quad w_2 = (1, 1, 1); \quad w_3 = (-1, 2, -1).$$

Этот базис обладает тем свойством, что любые его два вектора являются ортогональными, так как скалярное произведение $w_i \circ w_j = 0$ при $i \neq j$. Если сделать эти векторы единичной длины, то они будут иметь значения:

$$w_1 = \frac{1}{\sqrt{2}}(-1, 0, 1); \quad w_2 = \frac{1}{\sqrt{3}}(1, 1, 1); \quad w_3 = \frac{1}{\sqrt{6}}(-1, 2, -1).$$

Можно проиллюстрировать представление некоторого сигнала с помощью этого ортонормированного базиса.

Пусть мы имеем сигнал $S = (8, 16, 20)$. Тогда, используя стандартный базис, можем записать:

$$S \circ w_1 = \frac{1}{\sqrt{2}}(-8 + 9 + 20);$$

$$S \circ w_2 = \frac{1}{\sqrt{3}}(8 + 16 + 20);$$

$$S \circ w_3 = \frac{1}{\sqrt{6}}(-8 + 32 - 20).$$

Отсюда сигнал можно записать как

$$\begin{aligned} S &= (S \circ w_1)w_1 + (S \circ w_2)w_2 + (S \circ w_3)w_3 = \\ &= \frac{12}{\sqrt{2}}w_1 + \frac{44}{\sqrt{3}}w_2 + \frac{4}{\sqrt{6}}w_3. \end{aligned}$$

Тогда энергия будет равна

$$Q = d^2 = \frac{144}{2}w_1 + \frac{44^2}{3}w_2 + \frac{16}{6}w_3 = 72 + 645\frac{1}{3} + 2\frac{2}{3} = 720.$$

Энергию сигнала можно вычислить и непосредственно через квадрат длины вектора

$$Q = d^2 = 8^2 + 16^2 + 20^2 = 64 + 256 + 400 = 720.$$

Таким образом, используя ортонормированный базис, полную энергию можно получить в виде суммы энергетических составляющих, связанных с каждым базисным вектором.

Если учесть, что значение скалярного произведения двух векторов лежит в интервале $(-1, +1)$, то каждое конкретное значение этого произведения может служить мерой схожести векторов, так как оно равно $+1$, если векторы равны и -1 , когда они имеют противоположное направление.

3.4.2. Анализ на основе базиса Робертса

Для анализа полутоновых изображений используется базис Робертса, исследующий окрестности размером 2×2 и состоящий из четырех векторов $\{w_1, w_2, w_3, w_4\}$. Любая окрестность 2×2 действительных значений интенсивности может быть уникальным образом представлена в виде суммы

четырёх этих базисных векторов, взятых с соответствующими масштабными множителями. Относительная величина масштабного множителя характеризует сходство окрестности изображения с базисным вектором, поэтому может использоваться для определения структуры окрестности.

Ниже представляется базис Робертса:

$$w_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad w_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad w_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad w_4 = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

С помощью этого базиса можно представлять различные окрестности изображения размером 2×2 в виде произведения константного базисного изображения на число.

Например, область постоянных значений интерпретируется следующим образом:

$$\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} = 20/2 \left(\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right) = 10w_1 + 0w_2 + 0w_3 + 0w_4.$$

Примеры ступенчатых краев вертикально и диагонально ориентированных имеют такую интерпретацию:

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} = 0w_1 + \frac{2}{\sqrt{2}} w_2 + \frac{-2}{\sqrt{2}} w_3 + 0w_4.$$

$$\begin{bmatrix} +1 & +1 \\ -3 & +1 \end{bmatrix} = 0w_1 + \frac{4}{\sqrt{2}} w_2 + 0w_3 + \frac{-4}{2} w_4.$$

И пример для линии:

$$\begin{bmatrix} 0 & 8 \\ 8 & 0 \end{bmatrix} = 8w_1 + 0w_2 + 0w_3 + 8w_4.$$

3.4.3. Анализ на основе базиса Фрея – Чена

Данный базис предназначен для обработки окрестности изображений размером 3×3 . Стандартный базис для окрестностей размером 3×3 содержит 9 векторов, каждый из которых имеет одну единицу и 8 нулей следующего вида (рис. 11):

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Рис. 11. Стандартный базис Фрея – Чена

С помощью этого стандартного базиса можно представлять любую матрицу 3×3 в виде суммы соответствующих векторов стандартного базиса, умноженных на масштабный множитель, например (рис. 12):

1	7	0
0	0	0
0	4	0

1	0	0
0	0	0
0	0	0

0	1	0
0	0	0
0	0	0

0	0	0
0	0	0
0	1	0

Рис. 12. Пример представления матрицы в виде суммы векторов стандартного базиса Фрея – Чена

Однако разложение по стандартному базису не дает представления о структуре двумерной окрестности. Для анализа двумерной структуры более подходит базис Фрея – Чена, который состоит из множества ортонормированных масок.

Представление окрестности изображения с помощью базиса Фрея – Чена позволяет интерпретировать энергию изображения, рассматривая ее составляющие, соответствующие маскам градиента, пульсации, линии и т.п. Значение энергетической составляющей определяется степенью соответствия структуры изображения некоторому базисному вектору (маске).

Все векторы базиса Фрея – Чена разработаны специальным образом и имеют следующую структуру:

Градиент: $W_1 = 1/\sqrt{8}$

1	$\sqrt{2}$	1
0	0	0
-1	$-\sqrt{2}$	-1

 $W_2 = 1/\sqrt{8}$

1	0	-1
$\sqrt{2}$	0	$-\sqrt{2}$
1	0	-1

Пульсации $W_3 = 1/\sqrt{8}$

0	-1	$\sqrt{2}$
1	0	-1
$-\sqrt{2}$	1	0

 $W_4 = 1/\sqrt{8}$

$\sqrt{2}$	-1	0
-1	0	1
0	1	$-\sqrt{2}$

Линия $W_5 = 1/2$

0	1	0
-1	0	-1
0	1	0

 $W_6 = 1/2$

-1	0	1
0	0	0
1	0	-1

Лапласиан $W_7 = 1/6$

1	-2	1
-2	4	-2
1	-2	1

$W_8 = 1/6$

-2	1	-2
1	4	1
-2	1	-2

Константа $W_9 = 1/3$

1	1	1
1	1	1
1	1	1

Ниже приводится пример представления полутоновой окрестности с использованием базиса Фрея – Чена. Окрестность представлена матрицей следующего вида:

$$M = \begin{bmatrix} 10 & 10 & 5 \\ 10 & 5 & 5 \\ 10 & 10 & 5 \end{bmatrix}$$

Можно определить каждую компоненту для окрестности в направлении каждого базисного вектора. Скалярные произведения окрестности на базисные векторы и энергии компонент приведены в таблице 1:

Таблица 1

i	$N \circ W_i$	Q
1	0	0
2	$\frac{10+5\sqrt{2}}{\sqrt{8}} \approx 6$	36,5
3	$\frac{5-5\sqrt{2}}{\sqrt{8}} \approx -0,73$	0,53
4	$\frac{-5+5\sqrt{2}}{\sqrt{8}} \approx 0,73$	0,53
5	2,5	6,25
6	0	0
7	3,3	11,1
8	-0,8	0,7
9	23,3	544,4

Поскольку базис является ортонормированным, то полная энергия изображения равна сумме энергий его компонент. В данном случае эта сумма составляет 600. Полную энергию можно вычислить иначе, как произведение $M \circ M = 600$. Основную энергетическую составляющую дает средний уровень интенсивности вектора W_9 . Наибольшая энергия

остальных базисных векторов приходится на градиентный вектор W_2 . Если бы задачей поиска было обнаружение градиентов направления вектора W_2 , то в этом случае можно говорить о нахождении такого признака.

Алгоритм обнаружения окрестностей, имеющих высокую энергию для заданных векторов W_i , можно построить следующим образом.

Исходное полутоновое изображение In анализируется на выявление энергии подпространства, заданного вектором S , превышающей пороговый уровень Q_p . Бинарное выходное изображение Out содержит значение 0 для пикселей с энергией окрестности N , ниже уровня шума и порогового уровня и 1 – в противном случае. Вектор S длиной 8 имеет для j -го элемента $S_j = 1$, если вектор W_j входит в заданное подпространство векторов, и -0 , в противном случае.

1. Цикл по i .
2. Цикл по j .
3. Если $In[i, j]$ граничный пиксел, то $Out[i, j] = 0$.
4. Вычисляется энергия $Q_0 = (N \circ W_0)^2$. Задается энергия подпространства $Q_s = 0$.
5. Цикл по k .
6. Если $S(k) = 1$, тогда $Q_s = Q_s + (N \circ W_k)^2$.
7. Конец цикла по k .
8. Если $Q_s < Q_{min}$, тогда $Out[i, j] = 0$.
9. Если $\frac{Q_s}{N \circ N - Q_0} < Q_p$, тогда $Out[i, j] = 0$.
10. $Out[i, j] = 1$.
11. Конец цикла по j .
12. Конец цикла по i .



4. Обработка цветных изображений

4.1. Физические свойства цвета

В задачах компьютерного зрения цвет играет основную роль, так как обработка цветных изображений по сравнению с полутоновыми изображениями обладает более широким спектром подходов, методов и возможностей их решения. В настоящее время имеется широкий набор аппаратных и программных возможностей по работе с цветными изображениями.

Электромагнитные излучения с длинами волн в диапазоне 400–800 нанометров представляют собой видимый свет. В этот диапазон входят все составляющие спектр света от фиолетового до красного. Белый свет получается смешением световых компонент видимого диапазона.

Кроме видимого диапазона имеется аппаратура детектирования излучения и за его пределами – ультрафиолетового (длина волн меньше 400 нм) и инфракрасного (длина волн больше 800 нм), что дает возможность обрабатывать и эту информацию в задачах компьютерного зрения.

При освещении объектов источником света возникает отраженный свет, который и воспринимается наблюдателем или датчиком камеры. При этом можно фиксировать три основных фактора:

- **энергетический спектр**, который дает распределение световой энергии по длинам волн;
- **отражательная способность** поверхности объекта как результат отраженного света по отношению к падающему;
- **спектральная чувствительность** датчика, на который падает отраженный свет.

Кроме указанных основных факторов, на восприятие сцены влияют и ряд других, связанных как с самими объектами, так и с датчиками, принимающими свет. Для объектов сцены – это материал, из которых они изготовлены, качество и обработка поверхности, а также взаимное влияние объектов на степень их освещенности. Факторы, связанные с датчиками, заключаются в основном с технологическими особенностями камер. Если говорить о восприятии сцены человеком, то следует учитывать особенности зрительной системы человека.

4.2. Цветовые модели

4.2.1. Цветовая модель RGB

В графических системах компьютера широко применяется трехцветная модель RGB (красный-зеленый-синий). Современные системы имеют основной трехбайтовый (24 бита) формат кодирования цвета, при котором на каждую компоненту цветности выделяется один байт памяти. Это дает возможность получать более 16 миллионов цветов. Кроме этого формата еще используются варианты 15 бит (по 5 разрядов на компоненту цвета, 32 768 цветов) и 16 бит (красный – 5 бит, зеленый – 6 бит, синий – 5 бит, 65 536 цветов).

Один байт памяти дает возможность получения 256 оттенков каждой компоненты цвета от 0 до 255. Коды (255, 0, 0), (0, 255, 0), (0, 0, 255) дают соответственно красный, зеленый, синий цвета. Цвета, у которых коды каждой компоненты равны между собой, имеют полутоновые оттенки, число которых равно 256. Для полутоновых изображений достаточно использовать формат кодирования размером в один байт.

Система RGB называется аддитивной цветовой системой, так как ее цвета получаются суммированием основных цветов с черным цветом (0, 0, 0). В обработке информации иногда используют нормировку цветовых компонент по следующим соотношениям:

- интенсивность $I = (R + G + B)/3$;
- цветовые компоненты $r = R/(R + G + B)$; $g = G/(R + G + B)$;
 $b = B/(R + G + B)$.

При этом для всех цветов сумма $r + g + b = 1$.

4.2.2. Цветовая модель CMY

Цветовая модель CMY используется для цветной печати на белой бумаге. Она носит название субтрактивной, и ее цвета получаются на основе вычитания основных цветов из белого цвета. Название системы образовано от английских слов Cyan-Magenta-Yellow (голубой-пурпурный-желтый). Эти цвета соответствуют печатным краскам, поглощающим цвет одного из основных цветов RGB:

голубой цвет = $RGB - R$ (голубая краска поглощает красный цвет);

пурпурный цвет = $RGB - G$ (пурпурная – зеленый цвет);

желтый цвет = $RGB - B$ (желтая – синий цвет).

В этой цветовой модели белый цвет кодируется как (0, 0, 0), а черный – (255, 255, 255). Соответственно, цвета голубой, пурпурный, желтый кодируются (255, 0, 0), (0, 255, 0), (0, 0, 255).

4.2.3. Цветовая модель HIS

Модель Тон-Насыщенность-Интенсивность (Hue-Saturation-Intensity) в своей кодировке имеет значение интенсивности света I и два значения цветности – тон H и насыщенность S . Система используется в графическом дизайне (графических редакторах) из-за удобства контроля яркости и тона, а также в некоторых алгоритмах компьютерного зрения. Последнее применение объясняется более тесной связью модели со свойствами поверхности объектов.

Тон H определяется как угол между вектором насыщенности и осью красного цвета пространственной модели цветовой системы, и его значение меняется от 0 до 2π . При этом красному цвету соответствует угол 0 , зеленому – $2\pi/3$, синему – $4\pi/3$. Насыщенность S характеризует чистоту цвета. Значение $S = 1$ соответствует чистому (насыщенному) цвету, а $S = 0$ – ненасыщенному цвету (оттенок серого). Интенсивность I лежит в пределах от 0 до 1 .

4.2.4. Цветовые системы YIQ и YUV

Телевизионный стандарт NTSC использует цветовую модель YIQ, содержащую компоненту яркости Y и две цветовых компоненты I и Q . Черно-белые телевизоры используют только компоненту яркости. Линейное преобразование цветовой модели RGB в модель YIQ выполняется по следующим соотношениям:

$$\begin{aligned} \text{Яркость} & Y = 0,30R + 0,59G + 0,11B; \\ \text{Красный-голубой} & I = 0,60R - 0,28G - 0,32B; \\ \text{Пурпурный-зеленый} & Q = 0,21R - 0,52G + 0,31B. \end{aligned}$$

Учитывая большую чувствительность зрения человека к яркости, чем к цвету при кодировании на компоненту Y выделяют больше разрядов, чем на I и Q .

В алгоритмах сжатия и в некоторых цифровых видеосистемах используется похожая на YIQ цветовая модель YUV. Преобразование цветовой системы RGB в цветовую модель YUV выполняется по соотношениям:

$$\begin{aligned} Y &= 0,30R + 0,59G + 0,11B; \\ U &= 0,493 (B - Y); \\ V &= 0,877 (R - Y). \end{aligned}$$

4.3. Гистограммы цветных изображений

Гистограмма служит для представления распределения цветов на изображении. С этой целью выполняется подсчет количества пикселей

каждого оттенка цвета и отображения их в виде столбчатой диаграммы. Подсчет пикселей ведется раздельно для каждой компоненты цвета.

В некоторых случаях используется построение приближенных гистограмм, которые требуют меньшего объема вычислений с приемлемой точностью результатов. Например, для 24-битного представления цвета используются 6-битовые коды, составленные из двух старших разрядов каждого байта цветов RGB. В этом случае гистограмма имеет всего лишь 64 значения пикселей.

Гистограммы используются в решении задач поиска в базах данных и решении задач распознавания. Для этого вводится понятие пересечения гистограмм $Int(I, M)$ тестируемого изображения и гистограммы-образца, которое определяется как сумма минимальных значений из каждой пары соответствующих пикселей двух указанных гистограмм:

$$Int(I, M) = \sum_{j=1}^K \min\{I(j), M(j)\},$$

где $I(j)$, $M(j)$ – j -е элементы гистограмм тестируемого изображения и образца соответственно.

Если величину пересечения гистограмм нормировать общим числом пикселей, то полученное значение можно использовать как меру сходства $S(I, M)$ гистограмм и, соответственно, их изображений:

$$S(I, M) = \frac{Int(I, M)}{\sum_{j=1}^K M(j)}.$$

4.4. Текстура

4.4.1. Понятие текстуры

Текстура играет важную роль в компьютерном зрении и распознавании. Она определяет многие свойства изображения, что оказывает существенное значение для классификации областей изображения.

Текстура дает пространственное распределение цветов и интенсивности на изображении. Анализ этих характеристик позволяет выделять и классифицировать содержание изображения, определять качество материала объектов – вода, почва, деревья, камень, песок облака и т. д.

Понятие текстуры можно определить с двух точек зрения:

1. **Структурный подход.** С точки зрения структуры текстура – это множество примитивных текселей, расположенных регулярным порядком или повторяющимся.

2. **Статистический подход.** В этом случае текстура характеризуется числовыми характеристиками распределения значений интенсивности на изображении.

Далее будут рассмотрены эти два варианта описания текстур более подробно.

4.4.2. Тексельное описание текстур

При тексельном описании текстуры дается описание текселей и их пространственных взаимосвязей. По методу Тусерьяна и Джейн в качестве текселей выбираются области изображения, выделяемые в результате обработки изображения на основе, например, пороговой бинаризации. А пространственная взаимосвязь текселей определяется построением для них диаграммы Вороного.

Диаграммы Вороного строятся так. Имеется множество текселей S . Для двух текселей P и Q из этого множества проводится перпендикуляр, делящий отрезок PQ пополам. Плоскость таким образом делится на две полуплоскости такие, что все точки одной полуплоскости ближе лежат к P , а другой – к Q . Полуплоскость, содержащую точки, более близкие к точке P по отношению к серединному перпендикуляру отрезка PQ обозначим через $H^Q(P)$. Такую операцию можно провести для всех точек Q из множества S . Для каждой точки таким образом строятся многоугольники Вороного, содержащие все точки, более близкие к ним, чем к любым другим из множества S . Например, для круговых текселей многоугольники имеют вид, показанный на рисунке 13.

После построения многоугольников Вороного можно вычислить признаки формы многоугольников и использовать их для группировки многоугольников в кластеры, которые будут задавать области изображения с однородной структурой.

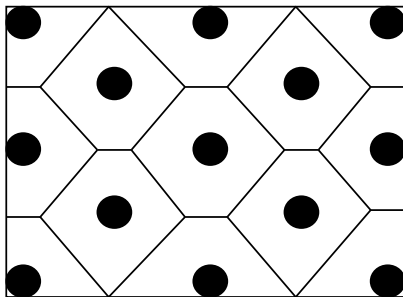


Рис. 13. Многоугольники Вороного для круговых текселей

4.4.3. Количественные характеристики текстур

При анализе естественных изображений выделение текстур является достаточно сложной задачей. Для них лучше подходит вычисление количественных характеристик, с помощью которых удобнее проводить обработку изображений.

Наиболее подходящей характеристикой в этом плане является обнаружение краев. Если области содержат фиксированное количество краев, то это может служить признаком для обнаружения заданных объектов. Кроме количества краев можно использовать и характеристику их направления.

Введем методику подсчета количественных характеристик текстур. Для некоторой области, содержащей N пикселей, выполняется обработка градиентным детектором краев, и для каждого пикселя p подсчитывается модуль (величина) градиента $Gr(p)$ и направление градиента $Dir(p)$. Затем подсчитывается плотность краев, как количество краев на единицу площади:

$$F = \frac{|\{p | Gr \geq T\}|}{N},$$

где T – некоторое пороговое значение градиента.

Набор характеристик можно дополнить свойствами ориентации текстур, используя нормированные гистограммы модуля и направления градиента для пикселей области R . Тогда количественное описание текстуры будет содержать соответствующие гистограммы:

$$F(h) = (H_{Gr}(R), H_{Dir}(R)).$$

Гистограммы обычно имеют небольшое количество разрядов (около 10), и их нормирование выполняется делением на размер области в пикселях. Для сравнения гистограмм двух изображений можно использовать параметр $L(H_1, H_2)$, называемый расстоянием между n -разрядными гистограммами:

$$L(H_1, H_2) = \sum_{i=1}^n |H_1(i) - H_2(i)|.$$

Для бинарных изображений в качестве гистограммы модуля градиента можно использовать двухразрядное представление: один разряд – темный пиксель, второй – светлый. А гистограммы градиента направления выбирать с числом разрядов, соответствующим числу выбранных направлений детектирования. Например, можно выбрать три направления – вертикальное, горизонтальное и диагональное.

Кроме гистограмм градиента модуля и градиента направления в качестве количественной характеристики текстуры используют гистограмму

интенсивности каждого пикселя по отношению к его соседним пикселям. Эта характеристика называется локальным двоичным разбиением. Вычисляется она для каждого пикселя в виде 8-битового значения. Каждый разряд b_i этой характеристики равен 0, если интенсивность i -го соседнего пикселя меньше или равна интенсивности самого пикселя, и -1 , в противном случае. В качестве меры сходства двух изображений также можно вычислять расстояние между их гистограммами, по формуле, аналогичной рассмотренной выше.

Еще одна используемая количественная характеристика для текстур – вычисление характерных признаков на основе матрицы вхождений, которая отображает некоторую пространственную структуру распределения значений пикселей на изображении.

Матрица вхождений представляет собой двумерный массив значений $S(i, j)$, которые указывают, сколько раз значение i встречалось в некотором пространственном отношении со значением j . Например, пространственное отношение может определяться так: «значение i является левым соседом для значения j ».

Ниже дается пример матрицы вхождений, в котором множеством допустимых значений для пикселей является множество значений яркости. Для каждой матрицы вхождений задается вектор пространственного отношения значений яркости. Этот вектор определяет взаимное положение двух пикселей, участвующих в вычислении характеристики. Индексы строк и столбцов составлены из всех вариантов значений яркости. Таким образом, подсчитывается количество вхождений в изображение двух яркостей со значениями (i, j) для пикселей, расположенных относительно друг друга на расстоянии dr по строке и dc по столбцу. Полутоновая матрица вхождений C_d для изображения I будет определяться следующим образом:

$$C_d(i, j) = |\{(r, c) | I(r, c) = i \text{ и } I(r + dr, c + dc) = j\}|,$$

где $d = (dr, dc)$ – вектор взаимного расположения пикселей.

Здесь под яркостью подразумеваются значение уровня полутонового изображения или цвет для цветного изображения.

Существует понятие стандартной полутоновой матрицы вхождений, которая может быть в двух вариантах – нормированной и симметричной. Нормированная матрица определяется по формуле

$$N_d(i, j) = \frac{C_d(i, j)}{\sum_i \sum_j C_d(i, j)}.$$

Симметричная полутоновая матрица вхождений имеет симметричный вид относительно главной диагонали, и ее значения вычисляются по матрице вхождений следующим образом:

$$S_d(i, j) = C_d(i, j) + C_{-d}(i, j).$$

На практике вместо матриц вхождений обычно используют набор характерных признаков, вычисленных на основе нормированной матрицы вхождений. Такой подход более эффективен с точки зрения объема вычислений. Такими характерными признаками являются энергия (Q), энтропия (P), контраст (K), однородность (D), корреляция (Cr):

$$Q = \sum_i \sum_j N_d^2(i, j);$$

$$P = - \sum_i \sum_j N_d(i, j) \log_2 N_d(i, j);$$

$$K = \sum_i \sum_j (i - j)^2 N_d(i, j);$$

$$D = \sum_i \sum_j \frac{N_d(i, j)}{1 + |i - j|};$$

$$Cr = \frac{\sum_i \sum_j (i - \mu_i)(j - \mu_j) N_d(i, j)}{\sigma_i \sigma_j},$$

где μ_i, μ_j – средние значения;

σ_i, σ_j – среднеквадратичные ожидания сумм $N_d(i)$, $N_d(j)$, вычисленных в направлении строк и столбцов изображения:

$$N_d(i) = \sum_j N_d(i, j);$$

$$N_d(j) = \sum_i N_d(i, j).$$

Наряду с перечисленными числовыми характеристиками в некоторых случаях для анализа текстур используют энергетические текстурные характеристики, а для обнаружения повторяющихся образов и описания зернистости текстурных элементов – вычисления автокорреляционной функции и спектр мощности [1].

Энергетические текстурные характеристики Лавса служат для оценки изменения содержания текстуры в пределах окна фиксированного разме-

ра. Для этого используется набор из девяти масок размерами 5×5. Результаты вычислений энергетических характеристик представляются для каждого пикселя вектором из 9 компонент. Для вычисления масок используются следующие векторы:

$$L5 = [1 \ 4 \ 6 \ 4 \ 1];$$

$$E5 = [-1 \ -2 \ 0 \ 2 \ 1];$$

$$S5 = [-1 \ 0 \ 2 \ 0 \ -1];$$

$$R5 = [1 \ -4 \ 6 \ -4 \ 1].$$



Названия векторов определяются их назначением – для вычисления среднего уровня (Level), выявления краев (Edge), пятен (Spot), ряби (Ripple). С помощью этих одномерных векторов получают двумерные маски перемножением пары одномерных масок из указанного выше набора. Одна из них используется как вектор-столбец, а вторая – как вектор-строка. Количество таких масок равно 16.

Перед расчетом энергетических характеристик выполняется предварительная обработка изображения путем вычитания из пикселей среднего локального значения интенсивности, в результате чего средняя интенсивность каждой пиксельной окрестности будет иметь значения, близкие к нулю. Окно обработки изображения для естественных сцен выбирается размером 15x15.

Затем в результате применения к обработанному изображению масок получают 16 отфильтрованных изображений F , по которым строятся энергетические текстурные карты E :

$$E_k(r, c) = \sum_{j=c-7}^{c+7} \sum_{i=r-7}^{r+7} |F_k(i, j)|,$$

где k – порядковый номер фильтра.

Из полученных 16 карт путем комбинирования симметричных пар типа L5S5/S5L5 получают девять энергетических карт: L5E5, L5S5, L5R5, E5E5, E5S5, E5R5, S5R5, S5S5, R5R5. Эти карты можно объединить в один трехмерный массив, в котором каждому пикселю в пределах окна 5x5 соответствует вектор из 9 текстурных атрибутов.

4.4.4. Текстурная сегментация

Текстурные характеристики могут использоваться для поиска в базах данных, задачах распознавания, а также для сегментации изображения на области с похожей текстурой. В последнем случае используются два подхода – анализ областей и анализ границ. Первый подход заключается в поиске группы или кластера пикселей, имеющих близкие текстурные

характеристики. При втором подходе решается задача обнаружения краев в текстурах и анализа параметров этих краев.

Для сегментации изображений, кроме прочих методов, широко используются энергетические текстурные характеристики. В общем случае сегментация естественных сцен в настоящее время остается нерешенной задачей.



5. Сегментация изображений

5.1. Понятие задачи сегментации

Суть сегментации изображений заключается в разбиении его на множество областей, покрывающих все изображение. При этом необходимо выделять такие области, которые представляют интерес в соответствии с решаемой задачей. В связи с этим форма и размеры областей могут иметь разный характер – линии, окружности, прямоугольники, многоугольники и т. п.

В процессе сегментации решаются две задачи. Первая – сегментация изображения на области с последующим анализом полученного результата. Вторая – изменение формы описания изображения, которое выполняется с учетом полученного разбиения.

Следует отметить, что не существует универсального метода решения задачи сегментации. Как правило, необходимо выбирать некоторый алгоритм из множества существующих решений для конкретной задачи. Возможно также использование нескольких алгоритмов с последующим выбором окончательного решения. В некоторых случаях можно разрабатывать алгоритм, комбинируя известные подходы.

5.2. Обнаружение областей

Можно сформулировать некоторые требования к выделенным областям, которым они должны удовлетворять. Например, следующие:

- области должны быть связными и однородными по какому-либо признаку – интенсивность, цвет, текстура;
- области должны иметь простую форму и не содержать внутри мелких отверстий и других деталей;
- смежные области сегментированного изображения должны существенно различаться по значению характерного признака;
- границы областей должны быть гладкими, без разрывов и зазубрин.

Перечисленные требования понятны, но не приходится надеяться на то, что они могут быть выполнены одновременно. В каждом конкретном случае можно пополнить список перечисленных требований, выполнение которых будет существенно облегчать решение задачи.

Кластеризация в задачах распознавания представляет собой процесс разбиения множества векторов признаков на подмножества, называемые

кластерами. Существуют разнообразные алгоритмы кластеризации. Ниже будут рассмотрены следующие методы кластеризации:

- классические алгоритмы;
- простые гистограммные методы;
- рекурсивный гистограммный метод Оландера;
- метод графового разбиения Ши.

5.3. Методы кластеризации

5.3.1. Классические алгоритмы кластеризации

Задача кластеризации состоит в разбиении векторов на подмножества векторов. В этих алгоритмах векторами обычно являются пиксели или окрестности пикселей. Множество векторов может содержать следующие компоненты:

- значения интенсивности;
- коды цвета или цветовые характеристики;
- вычисленные характерные признаки;
- текстурные характеристики.

Кластеризацию можно выполнять по любому характерному признаку. После кластеризации областей с помощью алгоритмов маркировки связанных компонент можно легко найти связанные области. Ошибка кластеризации D может быть вычислена по формуле

$$D = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - m_k\|^2,$$

где C_1, C_2, \dots, C_k – кластеры;

m_1, m_2, \dots, m_k – математические ожидания.

Эта величина показывает близость исходных данных к полученной кластеризации. Поиск оптимального решения путем перебора всех вариантов кластеризации с подсчетом величины D позволяет найти хорошее решение. Но этот путь требует большого объема вычислений. Поэтому на практике используют приближенные методы.

5.3.2. Итерационная кластеризация по математическому ожиданию

Решение задачи строится итерационным методом поиска экстремума. На первой итерации ($it = 1$) случайным образом задается K значений математических ожиданий для кластеров:

$$m_1(1), m_2(1), \dots, m_k(1).$$

Затем для каждого вектора x_i вычисляется расстояние

$$D = \sum_{k=1} \sum_{x_i \in C_k} \|x_i - m_k(it)\|^2.$$

Расчеты расстояния выполняются для всех кластеров. По результатам расчета вектор x_i перемещается в кластер C_j с ближайшим вектором математического ожидания. После этого выполняется переход на новую итерацию ($it = it + 1$) с уточнением значений математического ожидания и расчетом новых расстояний D .

Итерации повторяются до тех пор, пока для всех кластеров будет выполняться условие

$$C_k(it) = C_k(it+1),$$

т. е. кластеры остаются неизменными при переходе на очередную итерацию.

Алгоритм можно начинать не с задания случайных значений математических ожиданий, а со случайного разбиения кластеров и расчета для них математических ожиданий. Данный алгоритм в общем случае дает локально оптимальный результат. Он хорошо сходится и характеризуется умеренным объемом вычислений.

5.3.3. Алгоритм кластеризации ISODATA

Алгоритм ISODATA является разновидностью итерационных алгоритмов, в основе которого лежит метод разделения-объединения кластеров. В этом методе используется ковариационная матрица кластера вида:

$$\Sigma_k = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix},$$

где $\sigma_{ii} = \sigma_i^2$ дисперсия i -й векторной компоненты, а $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ представляет собой ковариацию i -й и j -й векторных компонент;

ρ_{ij} – коэффициент корреляции между i -й и j -й компонентами;

σ_i – среднееквадратичное отклонение i -й компоненты;

σ_j – среднееквадратичное отклонение j -й компоненты.

Алгоритм формирования кластеров строится следующим образом. Имеется множество кластеров C_1, C_2, \dots, C_k с математическими ожиданиями m_1, m_2, \dots, m_k и ковариационная матрица Σ_k для k -го кластера. Шаги алгоритма:

1. Поместить вектор x_i в кластер l , для которого достигается минимум

$$D_M = (x_i - m_l)' \Sigma_l^{-1} (x_i - m_l).$$

2. Объединить кластеры i и j , если

$$|m_i - m_j| < \tau_v,$$

где τ_v – пороговое значение дисперсии.

3. Разделить кластер k , если максимальное собственное значение Σ_k превышает порог τ_v .
4. Завершить работу, если для каждого кластера l выполняется условие

$$|m_i(t) - m_i(t+1)| < \varepsilon,$$

или если достигнуто максимально допустимое значение итераций.

5.3.4. Простые гистограммные методы

Гистограммные алгоритмы с точки зрения вычислений являются более эффективными, так как процедура вычислений является однопроходной.

В качестве примера ниже рассматривается алгоритм поиска моды гистограммы для кластеризации в пространстве измерений. Данный метод хорошо работает с изображениями, у которых однородные объекты образуют кластеры в пространстве измерений. В данном случае таким пространством измерений является гистограмма.

Сегментация изображений при этом выполняется путем обратного отображения кластеров из пространства измерений на изображения. Максимальные связные компоненты на гистограмме используются для выделения сегментов изображения. Для полутоновых изображений в пространстве измерений кластеризация может быть выполнена методом поиска точек минимума на гистограмме. В качестве кластеров выбираются интервалы значений между точками минимума. Маркировка кластеров на гистограмме выполняется индексами i . Все пиксели, попадающие в один кластер, маркируются соответствующим индексом.

Реализацию алгоритма можно строить на двух принципах. Первый состоит в поиске кластеров между точками минимума гистограммы. А второй – должен учитывать характеристики качества изображения в соответствии с решаемой задачей.

5.3.5. Рекурсивный гистограммный метод

Данный алгоритм Оландера является развитием гистограммного метода кластеризации. Идея алгоритма заключается в рекурсивной процедуре поиска мод. Первоначально поиск мод выполняют на гистограмме полного изображения. Затем этот поиск выполняют на гистограммах областей каждого сформированного кластера. Процедура выполняется, пока удастся производить дальнейшее разделение сегментированных областей.

Разбиение на кластеры выполнялось следующим образом. Сначала формировалась маска для выбора всех пикселей изображения. Затем строилась гистограмма изображения и выполнялось разбиение на кластеры в пространстве измерений. На основе сформированных кластеров устанавливалась принадлежность пикселей изображения найденным кластерам. Для каждого кластера выполнялась маркировка связанных компонент и строилось множество связанных областей, соответствующих кластерам. Затем для каждой связанной компоненты генерировалась новая маска и помещалась в стек.

Все маски в стеке предназначены для дальнейшей сегментации. В последующем каждая маска определяла пиксели для построения гистограммы. С помощью гистограммы выполнялось новое разбиение на кластеры в пространстве измерений. Процесс продолжается до исчерпания масок в стеке.

Существует разновидность метода Оландера, по которой гистограммы строятся не для компонент RGB, а для набора из трех переменных $(R+G+B)/3$, $(R-B)/2$ и $(2G-R-B)/4$. Такой алгоритм был предложен Ота в 1980 г.

5.3.6. Алгоритм графового разбиения Шу

Описанные выше алгоритмы Оландера и Ота хорошо работают на простых цветных изображениях с одноцветными областями, характерными для искусственных объектов и сцен. Естественные сцены и изображения обычно содержат много мелких деталей, сложную текстуру и большое число сегментированных областей. Для изображений с такими характеристиками был предложен алгоритм сегментации изображения на основе графовой модели, на которой решается задача разбиения графа [6].

Модель строится в виде взвешенного графа $G = (V, E)$, вершины которого представляют точки в пространстве измерений. Ребрам ставятся в соответствие веса $w(i, j)$, определяющие степень сходства вершин i и j .

Ставится задача разбиения графа на непересекающиеся множества V_1, V_2, \dots, V_m , которые имеют наибольшие сходства элементов внутри множеств и наименьшие сходства элементов разных множеств.

Если множество вершин V разбить на два непересекающихся подмножества A и B , то характеристика различия этих подмножеств может быть выражена как сумма весов ребер, соединяющих вершины из разных подмножеств. Удаление этих ребер приводит к разрезу (сечению) графа на два подграфа с вершинами A и B . Сечение графа определяется так:

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v).$$

С точки зрения решаемой задачи сегментации изображения ставится задача нахождения минимального разреза графа и повторение этой процедуры до тех пор, пока сегментированные области не станут достаточно однородными. В таком виде алгоритм может привести к формированию небольших множеств изолированных вершин, что является нежелательным при наличии больших однородных областей.

В связи с этим Ши предложил использовать другую характеристику – нормированного разреза. Для этого введена еще одна характеристика – ассоциация множества A и полного множества V :

$$asso(A, V) = \sum_{u \in A, t \in V} w(u, t).$$

Тогда нормированный разрез определяется следующим образом:

$$Ncut(A, B) = \frac{cut(A, B)}{asso(A, V)} + \frac{cut(A, B)}{asso(B, V)}.$$

Такая характеристика исключает возможность получения малых величин для небольших множеств изолированных вершин. Поэтому задача поиска минимального нормированного разреза в таком случае будет соответствовать решаемой задаче сегментации.

Характеристика связей вершин внутри множеств A и B определяется другой характеристикой – нормированной ассоциацией:

$$Nasso(A, B) = \frac{asso(A, A)}{asso(A, V)} + \frac{asso(B, B)}{asso(B, V)}.$$

Связь характеристик нормированного разреза и ассоциации связаны между собой следующим соотношением:

$$Ncut(A, B) = 2 - Nasso(A, B).$$

Данный алгоритм обладает неплохими результатами сегментации цветных изображений реальных сцен, но ввиду большого объема вычислений имеет ограниченное применение, особенно для систем реального времени.

5.4. Нарращивание областей

Нарращивание областей начинается с некоторой точки изображения путем присоединения соседних пикселей, не сильно отличающихся от пикселей области. Здесь рассматривается алгоритм, предложенный Харалик и Шапиро.

Сформированная область R из N пикселей характеризуется средним значением и разбросом:

$$\bar{X} = \frac{1}{N} \sum_{(r,c) \in R} I(r,c);$$
$$S^2 = \sum_{(r,c) \in R} (I(r,c) - \bar{X})^2.$$

Если все пиксели области R и пиксель y (кандидат на присоединение) независимы и их интенсивности описываются одинаковым нормальным распределением, то случайная величина

$$T = \left[\frac{(N-1)N}{(N+1)} (y - \bar{X})^2 / S^2 \right]^{\frac{1}{2}}$$

будет описываться t -распределением (распределением Стьюдента) с $(N-1)$ степенями свободы. Если значение T достаточно мало, то пиксель y присоединяется к области R . После этого выполняется пересчет среднего значения области и разброса.

Получение новых значений можно выполнять с учетом предыдущих значений по формулам:

$$\bar{X}_{new} = \frac{(NX_{old} + y)}{N + 1};$$
$$S_{new}^2 = S_{old}^2 + (y - \bar{X}_{new})^2 + N(\bar{X}_{new} - \bar{X}_{old})^2.$$

Если значение T велико, что пиксель y не присоединяется к области. Если значение y отличается от всех соседних областей, то данный пиксель берется за основу новой области.

5.5. Способы представления областей

5.5.1. Оверлейные представления

Данный метод представления областей используется для полутоновых изображений и заключается в выводе сегментированных областей поверх исходного изображения. Вывод границ областей выполняется цветными линиями или белым цветом, который хорошо контрастирует на изображении. При этом, как правило, используют толщину в несколько пикселей.

Во многих графических пакетах обработки изображений данный подход реализован в виде стандартной процедуры.

5.5.2. Маркированные изображения

Маркированные изображения используются в процессе обработки изображений как промежуточная процедура подготовки изображения для дальнейшей обработки. Суть операции заключается в индексировании областей уникальными идентификаторами. По исходному изображению строится новое, в котором каждая область изображения представляется в полутоновом виде и содержит присвоенный ей идентификатор. Для лучшего представления маркированного изображения используют методы выравнивания гистограмм изображения.

5.5.3. Кодирование границ

В некоторых случаях удобно представлять области их границами. В этом случае информация о сегментированных областях хранится отдельно от изображения, например, в виде линейного списка.

Примером такого рода линейного списка является цепной код Фримана. Принцип кодирования реализуется путем наложения границ областей на квадратную сетку, которая аппроксимирует кривую границы области. Для кодирования выбирается начальная точка, от которой идет переход по нужным направлениям к соседним точкам. Восемь возможных направлений кодируются от 0 до 7, им соответствуют углы от 0 до 315°. Разница между соседними углами составляет 45°.

Для привязки точек границы к системе координат достаточно хранить координаты начальной точки и сам цепной код. Такой способ представления областей требует достаточно небольших объемов памяти. При наличии внутри областей отверстий или других областей, они кодируются отдельно.

5.5.4. Квадрантные деревья

Квадрантные деревья представляют собой еще один способ изображения сегментированных областей. При этом хранится информация не только о границе, но обо всей области в целом. Каждый узел квадрантного дерева соответствует некоторой квадратной области и может иметь три состояния – заполненный (full, F), пустой (empty, E), частично заполненный (mixed, M).

Если узел помечен как заполненный, то все пиксели квадратной области принадлежат описываемой области. Если узел помечен как пустой, то все пиксели квадратной области не входят в область. Узлы, помеченные частично заполненными соответствуют квадратным областям, в которых могут быть пиксели, входящие в описываемые области и пиксели, не входящие в них. Дочерние узлы квадрантного дерева имеют только узлы, помеченные как частично заполненные. На рисунке 14 показан пример представления области изображения в виде квадрантного дерева. Обход областей выполняется по часовой стрелке, начиная с левого верхнего угла.

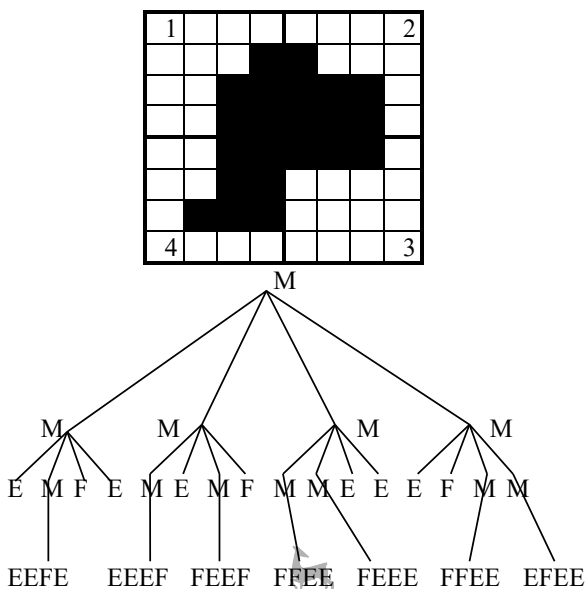


Рис. 14. Представление области изображения в виде квадрантного дерева (пример нумерации для первого уровня дерева показан в уголках области изображения)

Чем больше разрешающая способность изображения, тем большую глубину имеет квадрантное дерево. Способ представления областей в виде квадрантных деревьев широко используется в геоинформационных системах.

5.5.5. Таблицы свойств

Таблицы свойств представляют собой способ хранения информации областей в виде набора свойств: цвет, интенсивность, текстура, размеры, геометрические свойства и другие. В этом случае каждой области соответствует строка таблицы, а столбцы представляют свойства областей.

Таблицы свойств особенно удобны в решении задач поиска в базах данных. В этих случаях используются наборы свойств, позволяющие находить нужные изображения.

5.6. Обнаружение контуров

В некоторых задачах анализа изображений требуется работа не с выделенными областями, а с их границами, которые могут состоять из разных геометрических примитивов – отрезков прямых, дуг окружностей, эллипсов, произвольных кривых. Существуют различные алгоритмы обнаружения и выделения контуров. Ниже приводятся некоторые из них.

5.6.1. Прослеживание границ областей

Алгоритм прослеживания границ работает с изображениями, в которых выполнена сегментация областей или маркировка связанных компонент. Простое решение задачи заключается в том, что перебираются все точки изображения, и формируется список, содержащий первые граничные точки каждой области. Затем выполняется обработка каждой области, начиная с первой точки, путем поиска соседних пикселей границы. Поиск ведется по часовой стрелке до возвращения в исходную точку контура.

Описанный подход удобен для небольших изображений, которые располагаются полностью в памяти. Ниже рассматривается алгоритм, работающий с большими изображениями.

Алгоритм работает с предварительно маркированными областями изображения путем просмотра всех пикселей слева направо и сверху вниз. В результате обработки изображения на выходе получают границы областей, каждая из которых представлена в виде списка пикселей, в порядке обхода границ по часовой стрелке.

Предполагается, что фоновая область на изображении одна, и для тех областей, границы которых не входят в перечень обнаруживаемых, также присваивается метка фоновой области. Процесс организуется таким образом, что при последовательной обработке пикселей изображения, обнаруживаемый пиксель границы области относится к очередной точке границы соответствующей области. В каждый текущий момент обработки изображения в общем случае имеются границы областей, которые уже обработаны, частичные границы областей, находящихся в обработке, и границы областей, ожидающих обработки. Результаты работы алгоритма хранятся в виде структуры данных.

5.6.2. Обнаружение краев Кэнни

Алгоритм работает с полутоновыми изображениями и предназначен для выделения контуров. Детектор краев Кэнни генерирует тонкие фрагменты контуров изображения. Алгоритм оперирует с параметром сглаживания σ и двумя пороговыми значениями интенсивности.

Алгоритм обнаружения краев Кэнни описан в [1]. Сначала изображение сглаживается гауссовским фильтром с размахом σ . Затем для каждого пикселя сглаженного изображения вычисляется величина и направление градиента. Направление градиента используется для уточнения контуров за счет удаления пикселей, для которых величина градиента не превышает величины градиента в двух соседних пикселях по обе стороны от обрабатываемого вдоль направления градиента. Данный подход используется во многих алгоритмах для получения тонких контуров.

После обработки пикселей по величине градиента выполняется отслеживание непрерывных сегментов контуров. Анализ контуров начинается с краевых пикселей, у которых величина градиента превышает верхнее пороговое значение. Далее отслеживание контура выполняется на всех пикселях, значение градиента которых находится не меньше нижнего порогового значения. Обычно нижнее пороговое значение составляет половину величины верхнего порога.

5.6.3. Группировка контурных фрагментов

Если исходным изображением является изображение, содержащее маркированное контурное изображение с бинарными значениями пикселей, то задача отслеживания контуров становится более сложной, чем для маркированного изображения. В этом случае контурные линии не обязательно являются замкнутыми, и они могут образовывать сегменты с начальными, конечными пикселями, в виде сегментов-стыков и сегмен-

тов-углов. Алгоритм отслеживания контурных фрагментов должен выполнять операции:

- обнаружения нового сегмента;
- добавления нового пикселя к сегменту;
- завершение фрагмента (обнаружение конца);
- выявление стыка;
- выявление угла.

Результаты обработки формируются в виде структуры данных. Работа алгоритма заключается в анализе окрестности очередного пикселя. При этом определяется, является ли заданный пиксель изолированной точкой, начальной точкой нового сегмента, точкой стыка или углом. Если пиксель является концевой или внутренней точкой существующего сегмента, то алгоритм определяет и его идентификатор. Если он является угловой точкой или точкой стыка, то алгоритм определяет списки пикселей, входящих в эти сегменты и идентификаторов сегментов.

5.6.4. Обнаружение прямых и дуг окружностей

Определение прямых и кривых линий на полутоновых и цветных изображениях выполняют с помощью преобразования Хафа. В алгоритме используется массив, размерность которого определяется числом неизвестных параметров в уравнении семейства искомых кривых. При проверке каждого пикселя и его окрестности сначала определяется, присутствует ли в данном пикселе достаточно выраженный край. Если присутствует, то определяются параметры уравнения прямой $y = ax + b$, проходящей через данный пиксель.

После обработки всего изображения выполняется анализ полученных параметров. Для этого отыскиваются локальные максимумы в полученном массиве. Эти максимумы будут соответствовать параметрам наиболее вероятных прямых на изображении.

Аккумуляторный массив позволяет находить параметры бесконечно протяженных прямых и кривых, но он не дает возможность обнаружить отрезки прямых и сегменты кривых. Для этого случая требуется дополнительная детализация алгоритма, по которой выполняется анализ состава пикселей, образующих данный сегмент.

5.6.5. Обнаружение отрезков прямых и окружностей

Уравнение прямой $y = ax + b$ не подходит для обнаружения вертикальных прямых и отрезков. Более удобным является использование

уравнения прямой в виде $d = x \cos\theta + y \sin\theta$, где d – длина перпендикуляра к данной прямой, опущенного из начала координат, а θ – угол между этим перпендикуляром и горизонтальной осью. Если перевести это уравнение в экранную систему координат, то уравнение примет вид $d = x \cos\theta - y \sin\theta$. Экранная система координат ориентирована на дисплее таким образом, что начало системы координат находится в левом верхнем углу, горизонтальная ось координат направлена слева направо, а ось y – сверху вниз.

Аккумуляторный массив, который формируется в результате обработки изображения, содержит параметры d и θ , обнаруженных отрезков прямых. При этом вводится их дискретизация. Например, в алгоритме О'Гормана и Кловса [Шапиро], такая дискретизация составляла для d и θ значения 3 и 10° соответственно.

Преобразование Хафа может быть также использовано для обнаружения окружностей и других параметрических кривых.

Координаты точек окружности в экранной системе координат можно вычислять по формулам:

$$r = r_0 + d \sin\theta; c = c_0 - d \cos\theta;$$

где r, c – координаты точек окружности;

r_0, c_0 – координаты центра окружности;

d – радиус окружности;

θ – угол между радиусом, соединяющим точку окружности с ее центром и горизонтальной осью.

5.7. Подбор моделей сегментов

При обработке изображений и выделении контуров изображения может возникать задача замены выделенного контура на близкую ему математическую модель в виде прямой, дуги, окружности, эллипса и т. п. При таком подходе выполняется аппроксимация выделенного контура соответствующей моделью. Для подбора такой модели обычно используют метод наименьших квадратов. Например, при выявлении на изображении прямой или отрезка прямой с помощью преобразования Хафа [1], этот метод можно использовать для проверки пикселей, проголосовавших за данный элемент, на соответствие их уравнению прямой.

При использовании метода наименьших квадратов должен быть известен метод, с помощью которого выполняются подбор соответствующей модели из большого числа возможных вариантов. Ниже рассматривается более детально метод подбора прямой.

5.7.1. Подбор прямой

В качестве модели прямой можно взять уравнение с двумя параметрами:

$$y = f(x) = c_1 x + c_0.$$

Можно взять некоторое множество точек $((x_j, y_j), j = 1, n)$, выявленных как точки прямой, и проверить, лежат ли они на прямой. Для этого нужно подсчитать сумму квадратов отклонений точек от прямой:

$$S = \sum_{j=1}^n (f(x_j) - y_j)^2.$$

Из всех просчитанных моделей выбирается та, у которой достигается минимум суммы S .

Другой метод подбора заключается в использовании среднеквадратичной ошибки, которая определяется по формуле

$$RS = \left[\sum_{j=1}^n ((f(x_j) - y_j)^2) / n \right]^{1/2}.$$

В качестве критерия согласования модели с множеством точек может использоваться и критерий максимального отклонения:

$$M = \max(\{|(f(x_j) - y_j)|\}_{j=1, n}).$$

Следует отметить, что данный критерий дает локальную характеристику соответствия модели, которую применять можно не во всех случаях.

Аналитический расчет параметров модели прямой по методу наименьших квадратов можно выполнить следующим образом.

Ошибка рассогласования модели может быть вычислена как

$$\varepsilon(c_1, c_0) = \sum_{j=1}^n (c_1 x_j + c_0 - y_j)^2.$$

Тогда минимум этой функции находится в точке (c_1, c_0) , в которой выполняются равенства для частных производных:

$$\frac{\partial \varepsilon}{\partial c_1} = 0; \quad \frac{\partial \varepsilon}{\partial c_0} = 0.$$

Вычислив эти частные производные, получим:

$$\begin{aligned} \frac{\partial \varepsilon}{\partial c_1} &= \sum_{j=1}^n 2(c_1 x_j + c_0 - y_j) x_j = 0; \\ \frac{\partial \varepsilon}{\partial c_1} &= 2 \left(\sum_{j=1}^n x_j^2 \right) c_1 + 2 \left(\sum_{j=1}^n x_j \right) c_0 - 2 \sum_{j=1}^n x_j y_j; \end{aligned}$$

$$\frac{\partial \mathcal{E}}{\partial c_0} = \sum_{j=1}^n 2(c_1 x_j + c_0 - y_j) = 0;$$

$$\frac{\partial \mathcal{E}}{\partial c_1} = 2 \left(\sum_{j=1}^n x_j \right) c_1 + 2 \sum_{j=1}^n c_0 - 2 \sum_{j=1}^n y_j;$$

В матричной форме этот результат можно записать в следующем виде:

$$\begin{bmatrix} \sum_{j=1}^n x_j^2 & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n x_j y_j \\ \sum_{j=1}^n y_j \end{bmatrix}.$$

Решив эту систему уравнений, можно найти параметры наилучшей аппроксимирующей прямой. Иногда после анализа полученного результата выявляются точки, лежащие в стороне от идеальной прямой. В этом случае для улучшения результата отбрасывают такие точки, являющимися «выбросами», и повторяют расчет модели. Приближение вновь полученной модели к идеальной прямой может оказаться лучше, так как на нее не влияют точки-выбросы.

5.7.2. Подбор других моделей сегментов

Кроме подбора модели прямых и отрезков прямых могут выполняться аналогичные приемы для кривых, плоскостей и иных сегментов. В простейших случаях предполагается, что известны вид модели и набор исходных точек. В более сложных случаях, когда контурные цепочки могут быть представлены сочленением набора различных форм, применяют сегментацию кривых с последующим поиском моделей.

Реальные изображения могут содержать контурные сегменты в виде лент, углов и более сложные конфигурации. В зависимости от постановки задачи могут использоваться описанные выше алгоритмы и их модификации.

6. Методы распознавания образов

6.1. Основные задачи распознавания образов

Под образом обычно понимается группа объектов, объединенных по некоторому набору признаков, обладающих характерными свойствами, позволяющими выявлять его представителей. Способ отнесения распознаваемого элемента к какому-либо образу называется решающим правилом. При этом основную роль отнесения элемента к тому или иному образу играет так называемая метрика. Под метрикой понимается расстояние между элементами универсального множества, определяемое по заданному способу. Чем меньше это расстояние, тем более похожими считаются распознаваемые объекты и их образы. В связи с этим эффективность решения задачи распознавания образов в сильной степени зависит от выбора признаков образа и способа определения метрики [1].

При распознавании образов нередко применяются методы обучения, позволяющие повысить эффективность решения задачи распознавания. Обычно методами обучения называют процесс выработки реакции на группы внешних идентичных сигналов путем многократного воздействия на систему внешней корректировки. Такую внешнюю корректировку в обучении принято называть «поощрениями» и «наказаниями». Механизм генерации этой корректировки практически полностью определяет алгоритм обучения. Самообучение отличается от обучения отсутствием дополнительной информации системы о верности реакции.

В качестве наиболее известных примеров задач распознавания образов можно указать:

- распознавание объектов;
- распознавание текста;
- распознавание штрих-кодов;
- распознавание автомобильных номеров;
- идентификация лиц и других биометрических данных;
- распознавание речи.

Можно выделить три основных метода распознавания образов: методы перебора, методы с использованием характеристик образов и методы с использованием искусственных нейронных сетей.

В случае перебора выполняется простое сравнение объекта с базой данных, содержащей всевозможные модификации отображения. При этом база данных может формироваться из различных видов объекта. Например, могут использоваться изображения объекта, выполненные под различными углами, масштабами, смещениями, деформациями и так да-

лее. Для распознавания текста используются наборы гарнитур шрифтов с различными их начертаниями (полуужирный, подчеркнутый, курсив) и т. д. В случае распознавания звуковых образов происходит сравнение с некоторыми известными шаблонами (например, несколько вариантов слова, произнесенного одним лицом или разными людьми) ограниченно-го списка слов.

Метод, использующий наборы характеристик распознаваемого объекта, позволяет выполнять более глубокий анализ образа. В случае визуального распознавания это может быть определение различных геометрических размеров и частотных характеристик. Звуковой вариант распознавания в этом случае подвергается частотному, амплитудному анализу и т. д.

Использование искусственных нейронных сетей требует либо большого количества примеров задачи распознавания при обучении, либо специальной структуры нейронной сети, учитывающей специфику данной задачи. Тем не менее, его отличает более высокая эффективность и производительность.

Процесс решения задачи распознавания можно описать общей моделью классификации. Предполагается наличие множества из m известных классов, представленных набором примеров (эталонных образцов) объекта или в форме его описания. Кроме того, обычно предусматривается и особый класс объектов, состоящий из объектов, которые в процессе распознавания не были отнесены к какому-либо известному классу. Этот класс носит название «класс выбросов».

В процессе классификации объектов для каждого распознаваемого объекта выполняется назначение метки класса, которая определяет его принадлежность к конкретному классу. В качестве такого классификатора может выступать некоторое устройство или алгоритм.

Классификатор на входе имеет d -мерный вектор признаков объекта, который обрабатывается блоками, хранящими сведения о каждом из m классов объекта. Результаты обработки вектора признаков служат основанием для назначения соответствующих меток класса, включая и класс выбросов.

6.2. Ошибки распознавания

Качество работы классификатора можно охарактеризовать коэффициентом ошибок. Ошибкой считается отнесение входного объекта класса C_i к классу C_j , причем $i \neq j$ и C_j не является классом выбросов. Коэффициент ошибок определяется отношением числа ошибок в серии испытаний к общему числу испытаний. Эти коэффициенты могут различаться в


каждой серии испытаний, и для достаточного множества таких испытаний можно рассчитать математическое ожидание коэффициента ошибок данного классификатора.

Аналогично можно вычислить и коэффициент выбросов делением количества выбросов в серии испытаний на общее число испытаний.

Особо следует выделить случаи, когда результатом распознавания является отнесение объекта к одному из двух возможных классов по типу «Да/Нет». Например, возможны такие определения – хороший/плохой, присутствует/отсутствует, болен/здоров, есть ошибка/нет ошибки и т. д. Значение и цена ошибок отнесения к этим классам могут быть несимметричными. Например, если объект присутствует на изображении, то должна следовать определенная последовательность действий по обработке этой ситуации. В противоположном случае обработка ситуации будет другая. Ошибка, при которой дается ответ «Да» для случая «Нет», называется ложным срабатыванием или позитивной ошибкой. В противоположном случае, когда дается ответ «Нет» для случая «Да», ошибка называется ложным пропуском или негативной ошибкой.

Когда цена ошибки одного из вариантов значительно отличается от цены ошибки другого, то могут применяться дополнительные меры по снижению числа таких ошибок, в том числе за счет увеличения числа ошибок противоположного случая. Как правило, число ложных срабатываний увеличивается при попытке системы распознавания обнаружить большее число известных объектов.

Обычно с такими системами распознавания связывают понятия точности системы и полноты выборки распознаваемых объектов. Если обозначить через C_1 класс распознаваемых объектов, а через C_2 – класс объектов, не представляющих интерес, тогда точность системы t будет определяться как


$$t = \frac{k_1}{k_1 + k_2},$$

где k_1 – число найденных объектов класса C_1 ,

k_2 – число найденных объектов класса C_2 .

Полнота выборки определяется отношением числа найденных объектов k_1 класса C_1 к общему числу объектов K_1 этого класса:

$$p = \frac{k_1}{K_1}.$$

6.3. Распознавание по набору измерений

Для распознавания объектов необходимо иметь их описания. Такими описаниями может служить набор признаков измерений, характеризующих класс распознаваемых объектов. В набор признаков включаются:

- геометрические размеры объекта (площадь, ширина, высота, центр тяжести, периметр и т.п.);
- структура объекта (контур, число отверстий, краев, вытянутость, направление главной оси и т. д.);
- текстурные параметры;
- другие признаки.

Сформированный набор признаков представляется в виде упорядоченного вектора d признаков, каждый из которых представлен вещественным числом. На основании этого вектора можно ввести количественные оценки близости объекта к распознаваемому классу. Такая оценка формулируется в виде евклидова расстояния между двумя d -мерными векторами признаков x_1 и x_2 :

$$\|x_1 - x_2\| = \sqrt{\sum_{i=1,d} (x_1(i) - x_2(i))^2}.$$

6.4. Распознавание по элементарным признакам

Для распознавания объектов по набору элементарных признаков формируется соответствующий вектор признаков в результате обучения или на основе каких-либо моделей. При этом предполагается, что имеется m классов объектов, для каждого из которых заданы n_i эталонных векторов ($i=1, m$) размерностью d . Далее рассматриваются два метода распознавания с использованием эталонных образцов – по ближайшему среднему значению и по расстоянию до ближайших соседей.

6.4.1. Нахождение класса объекта по среднему значению

Классификация выполняется группировкой эталонных данных каждого класса по вектору математического ожидания класса:

$$m_i = \frac{1}{n_i} \sum_{j=1, n_i} x_{i,j},$$

где $x_{i,j}$ – j -й эталонный вектор признаков для класса i .

Распознаваемый объект с вектором признаков x_i относится к тому классу i , для которого разница $\|x_i - m_i\|$ наименьшая, по сравнению с другими классами. В некоторых случаях возникает необходимость отнесения распознаваемого объекта к классу выбросов, когда наименьшее евклидово расстояние превышает некоторый допустимый предел.

Данный метод хорошо работает для случаев компактного расположения векторов одного класса и при достаточном расстоянии между разными классами. Схема работы классификатора представлена на рисунке 15. Она классифицирует функции $f(x, K)$ через вычислительные операции над входным вектором признаков x с использованием некоторых сведений K , полученных на этапе обучения.

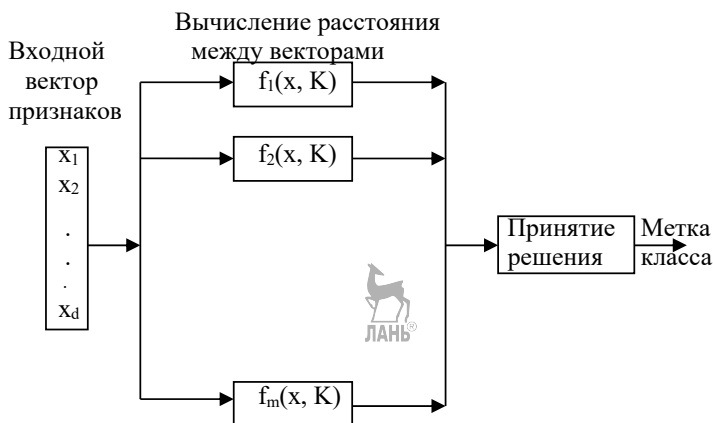


Рис. 15. Схема классификатора

Если классы недостаточно разнесены между собой, то возможно появление ошибок, связанных с тем, что отдельные векторы могут давать близкие результаты для разных классов. Еще больше ошибок может появляться при наличии сложной структуры классов: вытянутость точек класса, многомодальность (класс из нескольких кластеров), перекрытия областей и т. п. В этих случаях требуется модификация процедуры классификации. Например, вместо евклидова расстояния можно рассчитывать масштабированное евклидово расстояние:

$$\|x - x_c\| = \sqrt{\sum_{i=1,d} ((x(i) - x_c(i)) / \sigma_i)^2}.$$

Здесь масштабирование выполняется среднеквадратичным отклонением σ_i класса c по каждому элементу вектора признаков. Среднеквадратичное отклонение равно квадратному корню из дисперсии. Такое масштабирование необходимо во всех случаях, когда значения компонент вектора имеют разные единицы измерения или они сильно различаются по величине.

6.4.2. Нахождение класса по расстоянию до ближайших соседей

В данном методе отнесение распознаваемого объекта выполняется не по среднему значению класса, а по отдельному эталонному образцу, к которому он находится ближе всего. Данный подход может давать лучшие результаты и для случаев сложной структуры классов, и при наличии их перекрытия. При реализации этого метода вычисляются расстояния от исследуемого объекта до всех образцов базы данных и определяется минимальное расстояние. Данная технология является достаточно удобной в случаях изменения базы данных.

Принятие решения по отнесению объекта к определенному классу может приниматься не по одному минимальному значению расстояния, а в результате анализа нескольких наименьших значений. Если при этом находится несколько образцов одного класса, для которых расстояния наименьшие, то выполняется отнесение объекта к этому классу. При таком подходе требуется определить число k ближайших образцов, среди которых отыскивается решение. Если такого большинства не наблюдается, то объект относится к классу выбросов.

6.5. Структурные методы распознавания

В некоторых случаях оказывается недостаточным расчет числовых характеристик, когда они могут совпадать для разных классов. В таких случаях требуется привлечение дополнительных признаков распознавания. Подобные подходы носят название методов структурного распознавания образов.

В этих методах такими дополнительными признаками могут являться отношения между элементарными частями объекта. Эти отношения могут отображаться в виде графа, в котором вершинами выступают элементарные части объекта, а ребра между ними – суть бинарные отношения, определенные для них. Такие отношения могут быть не только бинарными, но и более высокого ранга – тернарными или кватернар-

ные. Описания таких отношений могут выполняться соответствующими гиперграфами.

Структурные методы полезны для распознавания сложных объектов, которые могут быть разбиты на объекты более низкого уровня.

Алгоритмы реализации структурных методов могут требовать дополнительных разработок в зависимости от области их применения.

6.6. Матрица неточностей

Матрица неточностей используется для представления результатов многочисленных экспериментов классификации объектов и оценки этих результатов. Матрица имеет размер $n \times m$, где n – число идентифицированных классов, включая и класс выбросов; m – число распознаваемых классов. На пересечении i строки j столбца указывается количество попыток классификации, в которых объект, принадлежащий классу i , был ошибочно идентифицирован как объект класса j .

Диагональные элементы этой матрицы, для которых $i = j$, содержат количество успешных экспериментов. В соответствии с этим идеально успешные эксперименты описываются матрицей неточностей, у которой все недиагональные элементы равны нулю, а диагональные – равны количеству экспериментов.

По матрице неточностей M_{ij} можно определить качество распознавания по каждому классу объектов. Для этого нужно вычислить коэффициенты ошибок. Например, коэффициент выбросов K_R определяется как частное от деления суммы элементов столбца выбросов на общее число экспериментов:

$$K_R = \frac{\sum R_i}{N},$$

где N – общее число экспериментов;

R_i – i -й элемент столбца выбросов.

Коэффициент общего количества ошибок K можно определить по следующему выражению:

$$K = \frac{\sum_{i \neq j} M_{ij}}{N}.$$

Аналогично можно определить коэффициенты ошибок по каждому классу объектов.

6.7. Дерево решений

Как правило, многие практические задачи распознавания образов требуют сравнений вектора признаков распознаваемого объекта с большим набором векторов признаков эталонных образов. Полное их сравнение приводит к выполнению значительных объемов вычислений. Это может оказаться технически трудновыполнимым и занимать слишком много времени. Кроме временных затрат могут требоваться и значительные материальные затраты, что во многих случаях может оказываться недопустимым.

В таких случаях удобным является метод построения дерева решений, который упрощает процедуру сравнения и уменьшает объем вычислений. Дерево представляет собой структуру, в которой узлы соответствуют некоторым признакам, а ветви – соответствующей классификации объекта на данном этапе.

Дерево решений разбивает все пространство поиска решений на части, которые выбираются на каждом этапе, исходя из анализа одного или нескольких признаков. Поиск завершается выходом на листовую узел. Сам результат распознавания определяется пройденным путем по дереву решений.

Частным случаем такого дерева решений является бинарное дерево, у которого в каждом узле только два варианта выбора дальнейшего анализа признаков, а выбор того или иного варианта осуществляется функцией выбора.

Если признаки выражаются числовыми значениями, то в качестве функции выбора используется порог, по отношению к которому производится выбор следующего узла классификации. Каждый узел бинарного дерева хранит только тип признака и пороговое значение, а листовые узлы содержат метку класса образа, к которому относится распознаваемый объект.

Подобные деревья решений проиллюстрированы на примере распознавания классов трехмерных объектов. В таблице 2 приведены объекты и вектора их признаков.

Таблица 2

Наименование признака класса	Номер признака класса	Классы									
		Параллелепипед	Тетраэдр	Октаэдр	Призма	Пирамида	Усеченная пирамида	Цилиндр	Конус	Шар	Тор
Количество вершин	1	8	4	6	6	5	8	0	1	0	0
Количество граней	2	6	4	8	4	5	6	2	1	0	0
Количество ребер	3	8	6	8	9	8	12	0	0	0	0
Количество окружностей в сечениях по осям X, Y, Z	4	0	0	0	0	0	0	1	1	3	6

Дерево решений приведено на рисунке 16. Узлы дерева соответствуют признакам классов. Они обозначены номерами этих признаков. Ветви, исходящие из узлов, соответствуют различным вариантам значений признаков. Ветви могут заканчиваться узлами других признаков или листовым узлом.

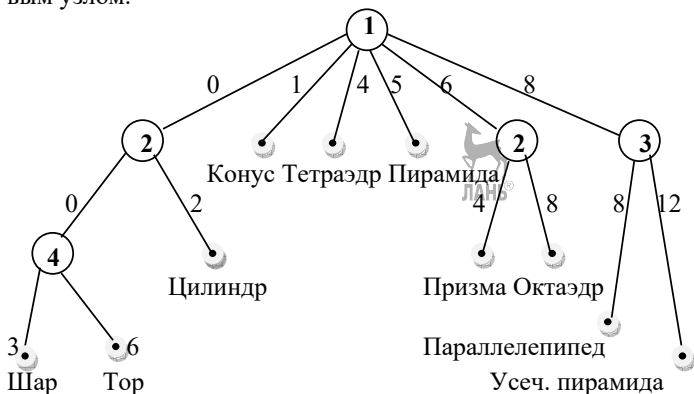


Рис. 16. Дерево решений

Аналогично построено бинарное дерево решений (рис. 17), в котором для каждого узла определен числовой порог для некоторого признака объекта.

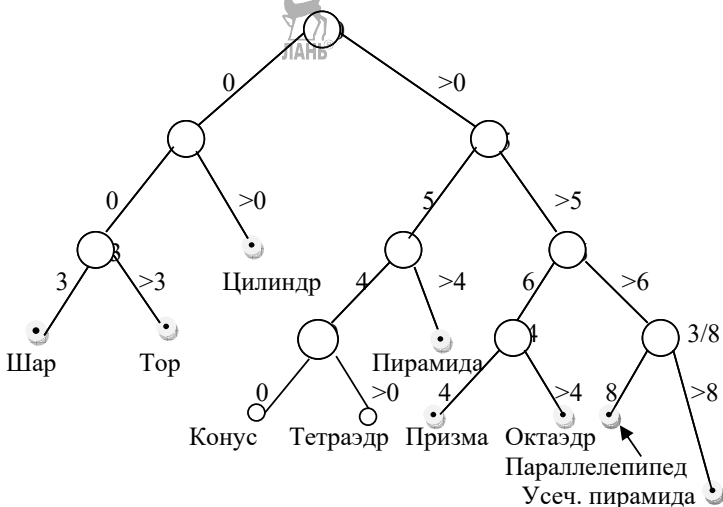


Рис. 17. Бинарное дерево решений

Если значение признака меньше или равно порогу, то выбирается левая ветвь дальнейшей классификации объекта, иначе – правая. Таким образом, из каждого узла возможен выбор одного из двух вариантов ветвления. Каждый узел бинарного дерева помечен номером признака и, через черту, значением порога. Поиск решения заканчивается листовым узлом, помеченным наименованием полученного класса. Деревья построены вручную в результате анализа данных для выбора подходящих признаков и пороговых значений.

В современных приложениях компьютерного зрения нередко требуются описанные решения для объектов, содержащих большие объемы признаков, и эталонных образцов, исчисляемых сотнями и тысячами. Поэтому ручное построение деревьев решений может оказаться затруднительным. Естественно, в таких случаях необходима автоматическая процедура создания дерева решений. Кроме того, для каждой задачи возможно построение нескольких деревьев, приводящих к решению. В связи с этим возможна дополнительная задача выбора оптимального дерева с точки зрения каких-то критериев. Наиболее вероятный критерий – построение дерева решения с минимальным числом уровней, что приводит к минимальному количеству проверок.

Вышесказанное можно проиллюстрировать на следующем примере. Пусть требуется распознать 4 класса со следующими эталонными образцами, приведенными в таблице 3.

По этим данным построены два бинарных дерева, различающиеся структурой и числом уровней проверок векторов признаков (рис. 18, 19).

Таблица 3

Класс	Признаки		
	А	Б	В
2	0	0	0
3	0	0	1
2	0	1	0
4	0	1	1
1	1	0	0
3	1	0	1
1	1	1	0
4	1	1	1

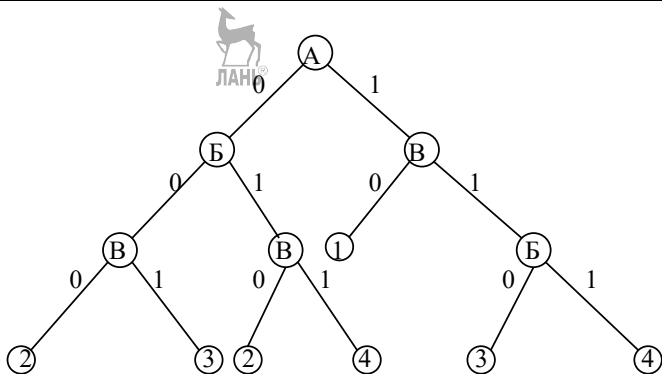


Рис. 18. Первый вариант бинарного дерева

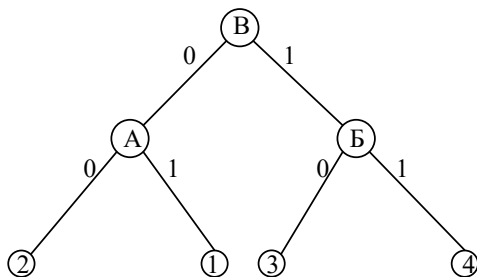


Рис. 19. Второй вариант бинарного дерева

6.8. Автоматическое построение дерева решений

Существует несколько методов автоматического построения дерева решений с разными определениями его оптимальности. Один из них основан на методах теории информации и использует понятие энтропии.

Энтропия множества событий $x = \{x_1, x_2, \dots, x_n\}$ определяется величиной $H(x)$:

$$H(x) = - \sum_{i=1}^n p_i \log_2 p_i,$$

где p_i – вероятность события x_i .

Для построения оптимального дерева решения используется понятие «прироста информации». Для этого вычисляется количество информации для событий класса по отношению к каждому его признаку.

Количество информации $I(C; F)$ переменной класса C с возможными значениями $\{c_1, c_2, \dots, c_m\}$ по отношению к переменной признака F с возможными значениями $\{f_1, f_2, \dots, f_d\}$ определяется следующим образом:

$$I(C; F) = \sum_i^m \sum_j^d P(C = c_i, F = f_j) \log_2 \frac{P(C=c_i, F=f_j)}{P(C=c_i)P(F=f_j)},$$

где $P(C = c_i)$ – вероятность появления класса C , имеющего значение c_i ; $P(F = f_j)$ – вероятность признака F , имеющего значение f_j , а $P(C = c_i, F = f_j)$ – совместная вероятность класса $C = c_i$ и переменной $F = f_j$. Эти априорные вероятности оцениваются по частоте событий в эталонных данных.

Проиллюстрируем автоматическое построение дерева решений с использованием понятия энтропии для предыдущего примера. Для этого подсчитаем вероятности и сведем их в таблицы 4, 5, 6.

Таблица 4

Класс (I)	1	2	3	4
$P(C = I)$	1/4	1/4	1/4	1/4

Таблица 5

Признак (J)	A	Б	В
$J = 0$	1/2	1/2	1/2
$J = 1$	1/2	1/2	1/2

Таблица 6

Классы	1	2	3	4
$A = 0$	0	1/4	1/8	1/8
$A = 1$	1/4	0	1/8	1/8
$B = 0$	1/8	1/8	1/4	0
$B = 1$	1/8	1/8	0	1/4
$V = 0$	1/4	1/4	0	0
$V = 1$	0	0	1/4	1/4

Теперь, вычислив меры количества информации, можно выбрать наилучший признак для выполнения проверки в корневом узле дерева.

$$\begin{aligned} (C; A) &= P(C = 1, A = 0) \log_2 \frac{P(C=1, A=0)}{P(C=1)P(A=0)} + \\ &+ P(C = 1, A = 1) \log_2 \frac{P(C=1, A=1)}{P(C=1)P(A=1)} + \\ &+ P(C = 2, A = 0) \log_2 \frac{P(C=2, A=0)}{P(C=2)P(A=0)} + \\ &+ P(C = 2, A = 1) \log_2 \frac{P(C=2, A=1)}{P(C=2)P(A=1)} + \end{aligned}$$

$$\begin{aligned}
& +P(C = 3, A = 0)\log_2 \frac{P(C=3,A=0)}{P(C=3)P(A=0)} + \\
& +P(C = 3, A = 1)\log_2 \frac{P(C=3,A=1)}{P(C=3)P(A=1)} + \\
& +P(C = 4, A = 0)\log_2 \frac{P(C=4,A=0)}{P(C=4)P(A=0)} + \\
& +P(C = 4, A = 1)\log_2 \frac{P(C=4,A=1)}{P(C=4)P(A=1)} = \\
& = 0 \cdot \frac{0}{1/4 \cdot 1/2} + 1/4 \cdot \frac{1/4}{1/4 \cdot 1/2} + 1/4 \cdot \frac{1/4}{1/4 \cdot 1/2} + 0 \cdot \frac{0}{1/4 \cdot 1/2} + \\
& + 1/8 \cdot \frac{1/8}{1/4 \cdot 1/2} + 1/8 \cdot \frac{1/8}{1/4 \cdot 1/2} + 1/8 \cdot \frac{8}{1/4 \cdot 1/2} + 1/8 \cdot \frac{1/8}{1/4 \cdot 1/2} = \\
& = 3/2.
\end{aligned}$$

Аналогичный расчет для двух других признаков дает
 $(C; B) = 3/2$; $(C; B) = 8$.

Таким образом, наилучшим признаком для корневого узла дерева решений является признак В, как имеющий наибольшую информацию для определения класса, что согласуется с ручным построением дерева. Подобное определение наилучшего признака для дальнейшего ветвления выполняется для всех узлов дерева решений.

6.9. Байесовский метод принятия решений

При этом подходе для принятия решения с наименьшим ожидаемым коэффициентом ошибок о принадлежности к классу используется априорная информация о распределении вероятности. Классификатор на основе наблюдаемых признаков относит объект к тому классу, к которому этот объект принадлежит с наибольшей вероятностью. Для решения задачи необходимо иметь для данных измеренных значений события x следующие распределения:

- условные вероятности $p(x/w_i)$ для каждого класса w_i ;
- априорную вероятность $P(w_i)$ для каждого класса w_i ;
- распределение безусловной вероятности $p(x)$.

В случае, если все классы w_i характеризуются собственными несовместимыми вероятностями, охватывающими все возможные случаи, то применимо правило Байеса для вычисления апостериорных вероятностей каждого класса по априорным вероятностям этих классов и распределениям условной вероятности для x .

Апостериорные вероятности для каждого класса определяются следующим образом:

$$P(w_i|x) = \frac{p(x|w_i)P(w_i)}{p(x)} = \frac{p(x|w_i)P(w_i)}{\sum_{j=1}^m p(x|w_j)P(w_j)}$$

Эти вероятности можно использовать для определения принадлежности объектов к одному из классов. Учитывая, что вероятность $p(x)$ одинакова для всех сравнений, то достаточно выполнять оценку принадлежности к определенному классу по максимальной величине $p(x/w_i) P(w_i)$.

На рисунке 20 даны классы w_1 и w_2 условных вероятностей значений x . Эти распределения перекрываются, что может приводить к ошибкам классификации объектов. Естественно, в данном случае установить порог для x , равный t , полагая, что при $x < t$ объект будет относиться к классу w_1 , а иначе – к классу w_2 .

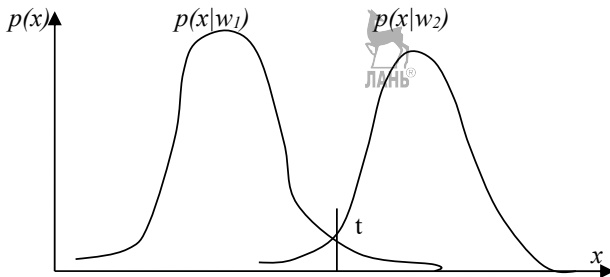


Рис. 20. Пример распределения вероятностей двух классов

Если при выделении класса w_1 значение $x < t$, а вероятность $p(x/w_1)$ приходится на область пересечения распределений двух классов, то возможно ложное срабатывание, в результате чего объект будет отнесен к классу w_2 . Если значение $x > t$, то в области пересечения распределений возможны ложные пропуски для класса w_1 .

6.10. Понятие методов кластеризации

Рассмотренные выше методы распознавания относятся к методам распознавания обучением с учителем. Для них характерно наличие структуры данных и желаемые результаты.

Методы распознавания без учителя иногда называют методами кластеризации. Для этих методов существует описание объектов в виде набора характеристик, называемых признаками. Кроме того, каждый объект описывается расстояниями до всех остальных объектов обучающей выборки. Разбиение выборки на группы схожих объектов позволяет упростить дальнейшую обработку данных и принятия решений, применяя к каждому кластеру свой метод анализа.

Имеется конечная обучающая выборка объектов. Требуется разбить её на непересекающиеся подмножества, называемые *кластерами*, так, чтобы каждый кластер состоял из объектов, близких по метрике, а объекты разных кластеров существенно отличались. При этом каждому объекту приписывается номер кластера.

Решение задачи кластеризации неоднозначно, так как не существует однозначно наилучшего критерия качества кластеризации. В связи с этим известен целый ряд эвристических алгоритмов, а также ряд алгоритмов, не имеющих чётко выраженного критерия, но осуществляющих достаточно разумную кластеризацию. Все они могут давать разные результаты, объясняемые тем, что число кластеров, как правило, неизвестно заранее и устанавливается в соответствии с некоторым субъективным критерием. Кроме того, результат кластеризации существенно зависит от метрики, выбор которой также субъективен и определяется экспертом.

6.11. Искусственные нейронные сети

Во многих задачах машинного зрения широко используются искусственные нейронные сети. Главная их особенность заключается в способности к обучению. Нейронные сети могут успешно применяться благодаря обучению сложной структуры образцов в многомерном пространстве с меньшими затратами памяти, чем это требуется для классификации по ближайшим соседям. Другая особенность заключается в возможности реализации параллельными алгоритмами, что дает существенное повышение производительности вычислительной среды.

Модель отдельного нейрона содержит тело нейрона, в котором выполняются вычисления. К нейрону подходят дендритические соединения, передающие входные сигналы $\{x_1, x_2, \dots, x_d\}$ от других нейронов или от датчиков. Тело нейрона умножает все входные сигналы на соответствующие коэффициенты $\{w_1, w_2, \dots, w_d\}$ и суммирует полученные произведения. Вычисленный результат преобразуется некоторой функцией f в выходной сигнал y :

$$y = f\left(\sum_{i=1}^d x[i]w[i]\right).$$

Выходной сигнал y посылается по аксону, разветвляющемуся на ряд дендритических соединений, которые передают его на другие нейроны в качестве входного сигнала. Примером простейшей функции f может являться пороговая функция

$$y = \begin{cases} 0, & \text{если } s \leq t; \\ 1, & \text{если } s > t, \end{cases}$$

где s – сумма произведений входных сигналов и коэффициентов.

Перцептрон представляет собой нейронную искусственную сеть, способную выполнять какую-то функцию и самообучаться. Алгоритм обучения начинается с выбора случайного набора весовых коэффициентов и пороговых значений. После просчета результата выполняется корректировка весового вектора в соответствии с положительным или отрицательным результатом для того или иного класса. В первом случае выполняется вычитание $g \cdot x$ из весового вектора, где g – масштабный множитель некоторой «выгоды», а во втором – прибавление этого же значения к весовому вектору. Эта операция проводится в цикле, который завершается, когда удастся правильно классифицировать образцы. В противном случае цикл повторяется с корректировкой масштабного множителя, например, уменьшение его вдвое. Предусматривается завершение цикла и по достижении максимально допустимого числа проходов.

Наряду с описанными искусственными нейронными сетями используются и нейронные сети прямого распространения, в которых каждый нейрон расположен в некотором слое. Сеть действует так, что нейрон получает входные сигналы из предыдущего слоя и передает выходной сигнал всем нейронам следующего слоя.

Самый нижний слой сети получает входные сигналы от датчиков, а выходные сигналы верхнего слоя являются результатом классификации. Промежуточные слои называются скрытыми. При этом отсутствует обратная связь между вышестоящими и нижестоящими слоями. Поэтому сеть называется сетью прямого распространения. Она работает по типу комбинационной схемы и не имеет памяти.

Сеть прямого распространения может обучаться настройкой весовых коэффициентов по последовательности эталонных образцов. В алгоритме обучения, называемом алгоритмом обратного распространения, ошибки классификации передаются от выходных слоев назад к входным слоям. Для обеспечения непрерывного управления обратной связью вход/выход используется сигмоидная функция, представляющая выходной сигнал в виде гладкой функции.

7. Поиск изображений на основе содержания

7.1. Задача поиска изображений

Достижения IT-технологий на современном этапе позволили оперировать с большими объемами данных и приемлемыми скоростями для оперативного использования. В настоящее время созданы разнообразные банки данных медицинских изображений, произведений искусства, спутниковых изображений и коллекций фотографий. Поэтому настоящей практикой востребована задача поиска нужных изображений.

Базы данных изображений могут содержать от сотен тысяч до десятков миллионов изображений. Чаще всего эти базы данных индексируются только по ключевым словам. Однако это бывает не всегда удобно и сужает возможности работы с изображениями. Существует немало задач, в которых необходим подбор изображений по отдельным элементам изображения – цвет, текстура, тип изображений, формы частей изображения и т. д. Именно такие элементы изображения входят в понятие их содержания. В настоящее время существуют некоторые коммерческие и исследовательские системы для поиска изображений в базе данных на основе содержания.

Для работы с базами изображений важным является формирование запроса на поиск изображений в соответствии с требованиями пользователя. Простой перебор изображений по некоторому критерию, как правило, приводит к большим объемам вычислений и, с этой точки зрения, является неоптимальным. Поэтому желателен способ поиска, обладающий какой-то эффективностью по сравнению с прямым перебором.

7.2. Поиск с использованием ключевых слов

Для облегчения задачи поиска многие организаторы баз изображений производят классификацию изображений путем назначения им ключевых слов и признаков их категорий. Ключевые слова и категории назначаются в виде текстовых атрибутов, присутствующих на самих изображениях. Такими атрибутами обычно являются: характер изображения, дата его создания, название источника или автора, названия ключевых объектов, имена и фамилии лиц и другие. Указание этих атрибутов позволяет выполнить индексацию изображений, после чего запросы поиска изображений могут быть сформулированы на языке запросов SQL, который в реляционных базах данных является стандартным языком запросов.

Кроме текстовых атрибутов на изображениях можно использовать некоторый атрибут для хранения набора значений, характеризующих изображения, которые также можно вводить в запрос в качестве ключевых слов. При этом увеличение набора характеристик приводит к более точному подбору изображений, соответствующих запросу пользователя.

Поиск с использованием ключевых слов связан с трудоемкой ручной работой назначения ключевых слов, а также с проблемой неоднозначности характеристик для разных пользователей. В связи с этим возникает необходимость поиска более эффективных методов поиска. Повышение эффективности поиска можно добиться за счет автоматического назначения ключевых слов или использования других методов поиска.

7.3. Поиск с использованием запросов по образцу

При данном подходе формируется некоторая табличная форма, содержащая набор значений, характеризующих изображение, и ограничения на них. Затем эта табличная форма используется для формирования запроса на языке SQL. В английской редакции запрос по образцу носит название *query-by-example* (QBE).

Примером системы подобного вида является Microsoft Access. В стандартных реляционных базах данных, имеющих текстовые или числовые атрибуты, организация запроса сводится к заполнению формы с параметрами образца и не требует дополнительных средств для реализации запроса.

Данный метод позволяет вместо набора данных использовать сами изображения, их эскизы или контуры искомых объектов. Система должна организовать поиск изображений, соответствующих заданному изображению. В этом состоит цель поиска изображений на основе содержания, и каждая система поиска предусматривает собственные средства задания запросов и определения сходства между запрашиваемым образцом и изображениями базы данных. При этом формируется набор возвращаемых изображений.

Таким образом, при организации запроса по образцу в качестве исходных данных могут использоваться фотографии, эскизы объектов или их контура, а также ограничения. В случае, когда изображение отсутствует, поиск выполняется только по ограничениям. Ограничения могут задаваться в виде ключевых слов, в виде задания объектов, которые должны быть на изображении, или описывая пространственные соотношения между объектами.

В общем случае запрос содержит цифровое изображение, сравниваемое с изображениями базы данных согласно применяемой мере расстояния между изображениями. Если расстояние равно 0, то считается, что изображение точно соответствует запросу. Изображение со значением отличным от 0 соответствует изображению в определенной степени. Степень соответствия снижается с ростом расстояния. Обычно поисковые системы выстраивают отобранные изображения по мере увеличения расстояния.

7.4. Меры расстояния между изображениями

Для оценки сходства изображений вводят количественные характеристики или меры расстояния. Сходство изображений обычно рассматривают по четырем группам: цвету, текстуре, форме и объектам. В соответствии с этим делением ниже рассматриваются подходы для расчета меры расстояния.

7.4.1. Характеристики цветового сходства

Характеристики цветового сходства рассчитываются разными способами. Например, можно указать процентное соотношение выбранных цветов и подбирать изображения, у которых аналогичное сочетание этих цветов. Данный подход достаточно простой, но применим он для изображений специфического характера. В общем случае такой подход будет давать изображения, далекие от содержания исходного, так как при этом не учитывается пространственное распределение этих цветов на изображении.

Другой способ основан на сопоставлении гистограмм изображений. Сравнение производится по K выбранным цветам. При этом используется оценка сходства цветов на двух изображениях. Если цвета совпадают или близки по значению друг к другу, то оценка приближается к 1, в противном случае она стремится к 0. Также учитывается и количество пикселей одного цвета на изображениях.

Еще один способ расчета меры расстояния использует цветовой макет. В этом случае в качестве исходного изображения используется сетка пикселей, каждому из которых пользователь задает цвет поиска из некоторой таблицы цветов. Допускается использование обычных изображений-образцов. Система выполняет поиск изображений, у которых пространственное распределение цветов близко к распределению цветов в образце.

Близость сходства определяется попиксельно, а мера расстояния сравниваемых изображений определяется в результате сравнения всех пикселей изображений:

$$d(I, Q) = \sum_g d(C^I(g), C^Q(g)),$$

где $d(I, Q)$ – мера расстояния между изображением I из базы данных и изображением-образцом Q ;

$C^I(g)$ – цвет клетки g изображения I ;

$C^Q(g)$ – цвет клетки g изображения Q ;

$d(C^I(g), C^Q(g))$ – мера расстояния между одноименными клетками изображений I и Q .

При этом цвет клеток может представляться разными способами. Например, обычно используются следующие:

- среднее значение цвета пикселя;
- среднее значение цвета пикселя и среднеквадратичное отклонение цвета;
- многозрядные цветовые гистограммы.

Мера расстояния, определяемая формулой расчета $d(I, Q)$, должна учитывать способ представления цвета клетки.

7.4.2. Характеристики текстурного сходства

Определение числовых характеристик для текстурного сходства требует более объемных вычислений. Изображения, определяемые как сходные, должны иметь одинаковое пространственное распределение цветов (или яркости), хотя сами цвета или яркости могут и не совпадать. Для оценки степени сходства текстур используют текстурные характеристики изображений, например, энергетические текстурные характеристики Лавса [7]. Иногда меры расстояния рассчитываются с учетом пространственного распределения текстуры и цвета изображений.

Для подсчета сходства изображений необходимо определить способ задания текстуры и метод расчета числовых характеристик меры расстояния, связанный с технологией задания текстур.

Рассмотрим сначала методы представления текстур. Одним из пространственных методов является задание текстуры с помощью вектора текстурного описания. Вектор состоит из набора числовых характеристик, соответствующих обобщенным параметрам текстуры изображения. Примерами таких векторов является пятикомпонентный вектор Харалика, содержащий признаки, извлеченные из матрицы вхождений, и вектор

с девятью компонентами, соответствующими энергетическим текстурным характеристикам Лавса.

Вектор текстурного описания может применяться для всего изображения. Однако эффективно данный подход работает только для изображений с однородной текстурой. Если изображения имеют иную текстуру, то векторы текстурного описания вычисляются для каждого пикселя отдельно. Для уменьшения объема вычислений окрестность пикселя для расчета вектора берется небольшого размера (например, 15x15). Затем производится сегментация изображения с присвоением меток для каждой области. По вектору текстурного описания и принадлежности пикселя к соответствующему кластеру рассчитывают текстурные меры расстояния.

Для расчета такой меры расстояния выбирается текстура на изображении и текстура образца. Затем выполняется сравнение векторов их текстурных описаний:

$$d_t(I, Q) = \min_{i \in I} \|T(i) - T(Q)\|^2,$$

где $T(i)$ – вектор текстурного описания пикселя i изображения I ;

$T(Q)$ – вектор текстурного описания для выбранного текстурного образца.

Если для каждого изображения процедурой кластеризации предварительно составить списки текстурных категорий, то можно уменьшить объем вычислений для определения меры расстояния. Дополнительного ускорения поиска можно добиться использованием индексации.

Использование меры расстояния от выбранного образца требует интерактивного указания текстуры пользователем, что затрудняет автоматическое её использование с указанным в запросе изображением. Наиболее удобно характеристику текстурного сходства можно сформировать путем обобщения на случай текстур характеристик на основе закрашенных сеток, описанных выше.

В этом случае на изображение, указанное в запросе, накладывается сетка, и для каждой клетки сетки вычисляется вектор текстурного описания. Аналогичная операция выполняется для изображений базы данных. Затем вычисляется расстояние на основе сеточного текстурного сходства

$$d_{gt}(I, Q) = \sum_g d_{tex}(T^I(g), T^Q(g)),$$

где d_{tex} – евклидово расстояние или иная мера расстояния между клетками сеток.

Аналогично расстоянию между цветовыми гистограммами можно определить расстояние между текстурными гистограммами.

7.4.3. Характеристики сходства формы

Выше рассматривались атрибуты изображений, которые характеризуют изображение в целом. В этом смысле они являются глобальными характеристиками. Атрибуты формы относятся только к отдельным объектам изображения, и задача поиска заключается в нахождении таких изображений, которые содержат объекты указанной формы. Для решения этой задачи требуются дополнительные операции обработки по идентификации областей. Обычно такая операция выполняется вручную, но в отдельных случаях возможна и автоматическая сегментация изображений.

Существует ряд свойств областей изображения [7], которые считаются глобальными характеристиками формы, так как они относятся к форме в целом. Сравнение форм выполняют разными методами статистического распознавания образов, а также структурными методами, при которых форма описывается перечислением примитивов и их пространственных отношений.

Обычно структурное описание выражается реляционным графом, поэтому сопоставление форм можно выполнять методами сопоставления графов. Эти методы представляют собой достаточно мощный аппарат решения задачи, но с увеличением числа элементов формы экспоненциально увеличивается объем вычислений. В связи с этим желательно иметь методы определения близости двух форм, может быть упрощенные, но работающие достаточно быстро.

При сопоставлении формы нередко требуется решать эту задачу с инвариантностью форм относительно смещения и масштаба. Иногда добавляется инвариантность и относительно вращения формы. В отдельных случаях инвариантность форм в отношении вращения не требуется ввиду возможных физических условий, по которым вращения невозможны. Например, здания, деревья и другие подобные объекты.

Существуют разные характеристики формы, обладающие различным качеством решения задачи и объемом вычислений. Ниже рассматриваются некоторые из них.

Гистограммные формы. Эти формы позволяют исключить из обработки области, которые не дают положительного сопоставления с формой образца. Метод привлекателен простотой вычисления расстояния между гистограммами двух форм. Главной задачей при таком подходе является определение переменной, по которой строится гистограмма.

Один из вариантов применения данного подхода может быть описан следующим образом. Предположим, что сравнение формы производится по связным областям бинарного изображения, которое имеет n строк и m

столбцов. Пиксели формы обозначены 1, а остальные – 0. Гистограмма ставит в соответствие каждой строке и каждому столбцу один разряд, значение которого определяется суммой 1 по строке и столбцу. Таким образом, ставится задача сравнения гистограмм с $n + m$ разрядами для областей одинакового размера.

При таком подходе можно добиться инвариантности относительно переноса и относительно небольших углов поворота и незначительных геометрических искажениях. В более общих случаях инвариантность относительно поворота может быть достигнута за счет вычисления осей наилучшего аппроксимирующего эллипса. Сопоставление форм в этом случае сопровождается сравнением больших и малых осей аппроксимирующего эллипса.

Еще один метод описания форм заключается в построении гистограммы по значению угла наклона касательной для граничных пикселей области. Этот подход дает инвариантность расчета относительно переноса и масштабирования, но не является инвариантным относительно поворота, так как угол наклона касательной вычисляется для области с фиксированной ориентацией.

Решение проблемы инвариантности относительно поворота возможно разными путями. Например, можно выполнить область, контролируя положение большой оси аппроксимирующего эллипса. Еще один подход предполагает циклический сдвиг гистограммы. Для K -разрядной гистограммы возможны K различных вариантов сдвига. В последнем случае можно отслеживать некорректные сдвиги гистограмм по сильному увеличению расстояния между гистограммами сравниваемых описаний формы.

Наконец, еще один способ заключается в нормировке гистограмм, при которой в качестве первого разряда выбирают разряд с наибольшим значением. Для улучшения результата, а также при наличии шумов, в качестве первого разряда гистограммы можно использовать несколько разрядов с наибольшими значениями.

Сопоставление границ. Данный метод связан с процедурами выделения и формирования описания границы области. Операцию выполняют для области запроса и области изображения из базы данных. Граница области может быть описана перечислением пикселей в виде последовательности символов, образующих границу области, или в виде многоугольника, аппроксимирующего границу области.

Классический метод сравнения описаний формы в виде последовательности символов основан на использовании рядов Фурье. Если границу можно описать аналитически в виде непрерывной функции, то эту

функцию можно разложить в ряд Фурье, а коэффициенты членов ряда использовать в качестве описания формы области.

Для дискретного случая граница представляется последовательностью точек $\langle V_0, V_1, \dots, V_{m-1} \rangle$, для которой определяются последовательности единичных векторов:

$$v_k = \frac{V_{k+1} - V_k}{|V_{k+1} - V_k|}$$

и последовательность кумулятивных разностей:

$$l_k = \sum_{i=1}^k |V_i - V_{i-1}|, k > 0, \\ l_0 = 0.$$

С помощью кумулятивных разностей рассчитываются коэффициенты ряда Фурье, которые можно использовать для определения меры расстояния для сравниваемых форм.

Обозначим через Q описание формы запроса, а через I – описание формы сравниваемой области изображения. Соответствующие последовательности коэффициентов Фурье формы запроса и области изображения из базы данных обозначим как $\{a_n^Q\}$ и $\{a_n^I\}$. Тогда меру расстояния для описаний форм, выполненных на основе коэффициентов Фурье, можно рассчитать следующим образом:

$$d_F(I, Q) = \left[\sum_{n=-M}^M |a_n^I - a_n^Q|^2 \right]^{\frac{1}{2}}.$$

Данный подход дает инвариантность меры расстояния только относительно переноса. Для получения инвариантности относительно других преобразований, поворота и масштабирования требуется меру расстояния использовать совместно с расчетом параметров масштабирования, поворота и начальной точки для поиска минимума значения $d_F(I, Q)$.

В случае представления областей формы запроса и изображений из базы данных аппроксимированным многоугольником описание формы может содержать длины сторон многоугольника и углы между его сторонами. Другим вариантом представления формы может служить множество координат вершин многоугольника и углов между его сторонами.

Пусть последовательность $Q = Q_1, Q_2, \dots, Q_n$ представляет собой описание формы области запроса в виде троек $Q_i = (X_i, Y_i, \alpha_i)$ – пара отрезков и угол между ними. Аналогичное описание выполняется для областей изображений из базы данных: $I = I_1, I_2, \dots, I_m$. При сравнении границ необходимо найти такое отображение Q на I , которое отображает отрезки прямых из описания формы запроса на отрезки похожей длины в описании формы области изображения. Сравнение выполняется поэлементно для отрезков и углов между ними.

Существует еще один метод «эластичного сопоставления» форм, при котором область из запроса деформируется таким образом, чтобы она стала максимально похожей на область изображения. Тогда расстояние между описанием формы из запроса и формы области изображения определяется по энергии, необходимой для деформации области запроса до наилучшего совпадения с областью изображения и характеристикой качества совпадения деформированной области запроса с областью изображения.

Эскизное сопоставление. Сопоставление форм выполняется по эскизному изображению, содержащему основные контуры изображения. Это изображение участвует в поиске изображений, содержащих подобные контуры.

Для поиска изображений в базе данных, близких к изображению запроса, выполняют предварительную обработку изображений базы данных, в результате которой получают промежуточные представления, называемые абстрактными изображениями. Операции предварительной обработки для получения абстрактных изображений состоят из следующих процедур.

Уменьшение изображения. Уменьшение изображения выполняют до заданных размеров, определяемых постановкой задачи, путем аффинных преобразований. Одновременно выполняют операцию по удалению шумов, как правило, используя медианный фильтр. Полученное изображение называется нормированным изображением.

Обнаружение перепадов. Обнаружение перепадов выполняют градиентным алгоритмом двумя шагами. На первом шаге определяют глобальные перепады по глобальному порогу. Величина глобального порога определяется через среднее значение и дисперсию градиента. На втором шаге определяют локальные перепады через локально вычисленные перепады. В результате этих операций получают очищенное контурное изображение.

Скелетизация и удаление избыточных контуров. В результате скелетизации и удаления избыточных контуров получают итоговое абстрактное изображение, содержащее очищенное от помех изображение контуров исходного изображения.

Эскизное изображение запроса тоже подвергается аналогичным операциям, описанным выше: нормировка размеров, бинаризация, скелетизация и удаление лишних контуров. Обработанное эскизное изображение называется линейным эскизом. В результате выполненных операций по обработке изображений переходят к сопоставлению линейного эскиза и

абстрактного изображения. Для сравнения обычно используется корреляционный алгоритм.

На линейный эскиз и абстрактное изображение накладывается сетка, и сравнение выполняется для соответствующих клеток. Для каждой клетки сетки изображения запроса вычисляется локальная корреляция с соответствующей клеткой сетки изображения из базы данных. Для повышения надежности результата сравнение выполняется несколько раз для нескольких значений сдвига положения клетки на изображении из базы данных. Из просчитанных значений сдвига выбирается то, которое имеет максимальное значение корреляции. Это значение является результатом сравнения двух клеток изображения линейного эскиза и абстрактного изображения. Характеристика сходства этих изображений определяется как сумма этих максимальных локальных корреляций, а мера расстояния обратно пропорциональна значению характеристики сходства:

$$d_{sk}(Q, I) = \frac{1}{\sum_g D_{cor}(\max\{A^I shift(g)\}, L^Q(g))'}$$

где $D_{cor}(A^I, L^Q)$ – корреляция между соответствующими клетками g линейного эскиза и абстрактного изображения;

$\max\{A^I shift(g)\}$ – максимальная корреляция из множества значений всех сдвигов клетки g .

7.4.4. Характеристики сходства объектов и их расположения на изображении

На практике задачи поиска изображений на основе содержания, как правило, ориентированы не на совокупность характеристик изображения (цвет, текстура, области и др.), а на смысловое содержание изображения – природа, цветы, деревья, картины, изображения людей, зданий и т. д. Ниже рассматриваются подходы поиска некоторых подобных классов изображений.

Обнаружение образов человеческих лиц. Обнаружение человеческих лиц является наиболее популярной задачей во многих областях применения компьютерного зрения. Эта задача достаточно сложная, так как оперирует с изображениями лиц разного размера, ракурса, освещенности, расположения, сочетания с другими предметами и т. п.

Для решения задачи поиска изображения человеческих лиц используют различные операции предварительной обработки изображений. Чаще всего такими операциями являются:

- перевод цветных изображений в полутоновые;
- нормировка освещенности;

-
- расширение диапазона яркостей выравниванием гистограмм;
 - фильтрация для устранения шумов;
 - выделение контуров и другие.

После обработки изображений используются различные приемы классификации изображений. Для этой цели часто используют искусственные нейронные сети с предварительным обучением. Результаты поиска могут быть существенно улучшены за счет введения различных ограничений, характерных для заданного класса изображений, а также за счет использования шаблонов для расположения глаз, носа и рта.

Обнаружение образов людей и человеческих тел. Решение этой задачи может быть выполнено с привлечением многих атрибутов по размерам, цвету, текстуре, контурам и т. д. При этом, как правило, выполняется и обработка изображения с получением набора характеристик для сопоставления изображений.

Интерес к этой задаче обусловлен широким спектром возможных применений, включающих в себя системы безопасности и ограничения доступа, человеко-машинный интерфейс, системы обеспечения безопасности на дорогах, средства обеспечения телеконференций и видеоконпрессии, исследования характера движения людей в различных ситуациях (в рабочей обстановке, на отдыхе, в спортивных соревнованиях) и т. д. [8].

Возможны различные подходы к методам обнаружения образов людей и человеческих тел. При этом может использоваться модель человека с нужными параметрами, а также поиск образа человека без модели. Неплохие результаты могут быть получены за счет использования информации от нескольких камер, которая позволяет строить пространственные модели.

Поиск человеческих тел может ставить своей задачей нахождения изображений, содержащих, например, открытые участки тела или участки определенной текстуры. Поиск образов людей без привлечения моделей обычно выполняется через обнаружение движения. Для этого все изображение разбивается на элементарные квадратные участки, на которых выделяются части, занимаемые человеческой фигурой. В каждом элементе разбиения изображения анализируются элементарные движения и формируется вектор перемещения. Специальным классификатором выделяется область, занимаемая человеческой фигурой на основе анализа векторов движения.

Другим подходом к решению задачи обнаружения человека является анализ статических кадров. При этом в качестве исходной информации для выделения областей возможного нахождения человеческой фигуры

вместо элементарных движений используются такие характеристики изображения, как границы, текстуры, результаты свертки изображения с банками фильтров и т. д. Например, в качестве элементарных признаков используются коэффициенты вейвлет-преобразования, полученные при обработке отдельных участков изображения. В процессе обучения на выборке изображений выделяется небольшая часть коэффициентов (наиболее общая), характеризующих целевой объект (человеческую фигуру), с учетом допустимой изменчивости в размерах и позах. На этапе поиска изображения человека используется скользящее окно переменного размера. Над частью сцены, попавшей в окно, выполняются операции нормализации, свертки с вейвлет-функцией и классификации. При этом классификатор принимает решение о присутствии человека в текущем окне.

Другим направлением обработки статических кадров является метод, использующий статистический анализ для построения моделей границ или всей фигуры человека. Для построения границ человеческой фигуры используются массивы изображений человеческих фигур с отмеченными на них характерными точками. Используя методы анализа главных компонент, получают компактное описание выбранной системы признаков, представленное системой векторов пониженной размерности, т. е. в исходном пространстве признаков формируется базис, в котором основная дисперсия направлена вдоль нескольких первых осей базиса. Представление модели человеческой фигуры в виде главных компонент кроме компактности обладает еще определенными обобщающими свойствами, позволяющими правильно классифицировать деформированные и искаженные фигуры.

При формировании обучающего множества для метода главных компонент вместо набора характерных точек также используются размытые бинарные образы, полученные в результате выделения силуэтов людей на изображениях сцен.

Алгоритмы обнаружения человеческой фигуры с заданной моделью человеческого тела используют априорное представление изображения человеческой фигуры в двумерной проекции для последующей сегментации, выделения фигуры и отдельных её частей. В связи с тем, что при произвольных движениях возникают проблемы, связанные с искажениями информации из-за взаимных перекрытий, изменений ракурса и частичным выходом за границы кадра, многие системы ограничивают допустимую изменчивость в пределах заранее сформулированной модели поведения человека в кадре. Предварительное выделение человеческой фигуры обычно осуществляется с помощью процедуры вычитания фона

в условиях стационарной или медленно изменяющейся фоновой обстановки и фиксированного положения камеры. Модель человеческой фигуры обычно представляется в виде проволочной модели с ленточным или дисковым покрытием.

Кроме двумерных моделей человеческого тела многие алгоритмы используют заранее сформированную 3D-модель человека в упрощенном виде. В качестве моделей используют проволочные или шарнирные модели, модели, составленные из цилиндров различных длин и диаметров, а также триангулированные модели поверхности тела.

Трехмерные модели позволяют более качественно определить положение, соответствующее данной плоской проекции. При выборе подходящей позы учитываются естественные ограничения, накладываемые человеческим телом на возможные положения частей тела. Для этого используют алгоритмы, учитывающие допустимые движения при установлении связанности в последовательности кадров. В качестве условия окончания поиска принимается состояние модели, имеющее минимальное отличие соответствующей двумерной проекции от исходного изображения.

Описанные методы обнаружения человеческой фигуры отличаются большими затратами вычислительных ресурсов и могут успешно реализовываться на базе графических процессоров за счет массового распараллеливания вычислительного процесса.

Обнаружение человеческой фигуры на стереоизображениях позволяет получить дополнительную информацию о пространственном расположении объектов на сцене и повысить точность выделения человеческой фигуры по сравнению с двумерными алгоритмами. Дополнительная информация о пространственной структуре сцены, во-первых, позволяет уменьшить зависимость от изменения освещенности сцены и упростить анализ изображений, связанный с учетом взаимного перекрытия объектов. Во-вторых, совместная обработка изображений, поступающих с нескольких камер, позволяет оценить реальное соотношение размеров обнаруженных объектов и, следовательно, обеспечить надежную классификацию. Разность между изображениями, полученными с левой и правой камер, позволяет построить карту разрывности, значения которой пропорциональны расстоянию соответствующих точек сцены от плоскости изображения. Для определения связанности объектов карта разрывности квантуется на несколько уровней, и элементы карты, принадлежащие одному уровню и имеющие пространственную близость, объединяются в один кластер (считается, что они принадлежат одному объекту). Сегментация объектов сцены, выполненная таким образом, не зависит от усло-

вий освещенности, текстуры поверхности, затенений и перекрытия объектов.

Пространственные отношения. После распознавания объектов на изображении можно определить их взаимное расположение. Параметры взаимного положения объектов на изображениях могут служить дополнительными характеристиками, позволяющими повысить качество решения задачи поиска изображений по запросу.

При решении задачи пространственного отношения объектов требуется и решение задачи отделения объектов сцены от фона. Решается она сегментацией изображения на базе цветовых и текстурных характеристик.

7.5. Организация базы данных

Проблема организации базы данных заключается в том, что объемы современных баз данных достаточно велики и их обработка может требовать значительных временных затрат. Поэтому стоит задача организовать базу данных таким образом, чтобы при обработке любого запроса достаточно было бы просматривать только часть базы данных. Достигается это обычно соответствующей организацией и индексацией базы данных. В настоящее время существует ряд стандартных методов индексации числовых, текстовых и пространственных данных. Кроме того, разрабатываются методы индексации изображений для использования их в системах поиска изображений на основе содержания.

7.5.1. Стандартные индексы

Наиболее известный способ организации базы данных заключается в указании некоторого атрибута, по которому необходимо построить индекс. В этом случае он является ключевым значением для каждой записи базы данных. Таким атрибутом могут являться фамилия, индивидуальный номер работника, для изображений – главный объект поиска, тип изображения, разрешающая способность и т. д.

В реляционных базах данных индекс используется системой для поиска по заданному значению атрибута. Этот подход позволяет быстро найти все множество записей базы данных, имеющих указанный атрибут с заданным значением.

Существует два типа индексов для поиска по заданному значению атрибута и по заданному интервалу, указанному в запросе. Первый тип носит название хэш-индексов, а второй – индексы на основе В-деревьев.

Хэш-индексы. Хэш-индексы используются для доступа к большому множеству записей базы данных с помощью хэш-таблиц. При этом значение ключа запроса преобразуется хэш-функцией в индекс (адрес) хэш-таблицы. Данный подход можно проиллюстрировать на примере базы данных, у которой ключевыми значениями являются числа. Примером простейшей хэш-функции в данном случае может служить, например, функция $f(x) = x \bmod N$. Значением $f(x)$ хэш-функции является остаток от деления значения ключа x на N , которое используется теперь в качестве номера записи для доступа к ней.

Если хэширование всех ключевых значений переводит их в собственные элементы таблицы, то время обращения к базе составляет одно и то же значение. Возможно, что несколько ключевых значений в результате хэширования дают указание на один и тот же элемент хэш-таблицы. Это приводит к так называемым коллизиям, которые требуют какого-то решения.

Таким образом, хэш-индексы позволяют просматривать только часть записей (данных), которые могут удовлетворять запросу, что значительно сокращает объем обработки базы данных. Хэш-индексы особенно эффективны для случаев конкретных значений ключей. Если в запросах используются интервалы значений, то эффективнее применять В-деревья.

Индексы на основе В⁺-деревьев. В-деревья и В⁺-деревья являются сбалансированными деревьями поиска и используются для индексации баз данных в случаях задания диапазона допустимых ключевых значений. В-деревья и В⁺-деревья различаются тем, что в В-деревьях значения ключей и данных хранятся как во внутренних, так и в листовых узлах, а В⁺-деревья хранят эти данные только в листовых узлах. Так как данные в базах данных должны храниться отдельно от индекса, то для поиска используют В⁺-деревья. Дерево поиска порядка p в каждом узле содержит не более $p-1$ ключевых значений и p указателей. Причем В⁺-деревья имеют разные форматы представления внутренних и листовых узлов.

Внутренние узлы В⁺-деревьев должны удовлетворять следующим требованиям:

- каждый внутренний узел дерева содержит указатели и ключевые значения в чередующемся порядке, причем крайними являются указатели; каждое ключевое значение K_i имеет указатель P_i слева на поддерево со значениями меньше или равны K_i , а справа указатель P_{i+1} на поддерево со значениями, превышающими K_i ;
- корневые узлы содержат не меньше 2 указателей на поддерева;
- некорневые узлы содержат не менее $\lceil p/2 \rceil$ указателей (наименьшее целое, большее $(p/2)$).

Листовые узлы V^+ -деревьев должны удовлетворять следующим требованиям:

- каждый листовой узел имеет формат $(K_1, Pr_1, K_2, Pr_2, \dots, K_{q-1}, Pr_{q-1}, P_{next})$; здесь K_i – ключевое значение, Pr_i – указатель на данные, а P_{next} – указатель на следующий листовой узел;
- указатель Pr_i указывает на запись или блок записей, имеющих значение ключа поиска K_i ;
- в каждом листовом узле хранится $\lfloor p/2 \rfloor$ значений (наибольшее целое, меньшее $(p/2)$);
- все листовые узлы располагаются на одном уровне дерева.

Поиск данных с помощью V^+ -деревьев при задании конкретного ключа или диапазона ключевых значений выполняется следующим образом. Система начинает обработку с корневого узла дерева, считывая его в память и выполняя двоичный поиск среди ключей, хранящихся в данном узле. Если узел содержит два смежных ключевых значения, между которыми находится значение, заданное поиском, то указатель между этими ключами указывает на поддерево, которое будет содержать искомое ключевое значение или наименьшее значение из указанного диапазона ключей. Если искомое ключевое значение меньше первого ключевое значения узла, то указатель слева от этого ключа ссылается на поддерево для дальнейшего поиска.

Аналогично, если искомое ключевое значение превышает последнее ключевое значение в узле, то указатель справа от этого ключа ссылается на поддерево для дальнейшего поиска.

После обнаружения подходящего поддерева поиск продолжается с корневого узла этого поддерева. Процедура выполняется рекурсивно до достижения листового узла.

Листовой узел обрабатывается двоичным поиском искомого ключевого значения или начального значения ключа K_i . Указатель P_i , связанный с этим ключевым значением, указывает на запись данных, содержащую это ключевое значение. Если требуется найти только одно ключевое значение, то процедура завершается возвратом найденной записи данных. При поиске же в рамках диапазона указатели P_{next} на следующие листовые узлы могут использоваться для нахождения оставшихся записей данных до обнаружения конца заданного в запросе диапазона ключей.

Поиск с помощью V^+ -деревьев широко применяется в реляционных системах управления базами данных. Они используются также в базах данных изображений для индексации отдельных числовых или текстовых

полей, сопоставленных с изображением. Приведенная индексация не предназначена для использования с многомерными данными.

7.5.2. Пространственная индексация

Данный вид индексации используется для баз с многомерными данными. Различные методы пространственной индексации связаны с соответствующей организацией структур данных. Наиболее популярен метод квадрантных деревьев с модификациями 2-мерных деревьев для бинарных изображений и K -мерными деревьями для k -мерных данных.

Квадрантные деревья представляют собой иерархические структуры четвертого порядка, которые разбивают пространство поиска двумерных данных на четыре квадранта на каждом уровне дерева. K -мерные деревья являются расширением бинарных деревьев поиска и предназначены для поиска среди k -мерных данных.

Еще один метод пространственной индексации основан на методе B -деревьев и рассчитан на обработку данных большой размерности и носит название R -деревьев. Как правило, метод R -деревьев используется для систем, работающих с пространственной информацией. Объекты данных индексируются с использованием n -мерного минимального описывающего прямоугольника, который сужает область просмотра данных. Каждому существующему объекту присваивается свой уникальный идентификатор. Форматы внутренних и листовых узлов дерева различны. В листовых узлах хранятся идентификаторы объектов данных, а во внутренних узлах дерева – записи, содержащие указатель на нижеследующий узел R -дерева, и охватывающий прямоугольник объектов, входящих в записи нижеследующих узлов. Существует несколько разновидностей R -деревьев, например, R^+ -деревья, R^* -деревья и др.

7.5.3. Индексация в системах поиска изображений на основе содержания

Рассмотренные выше методы индексации изображений ориентированы на случаи, когда мера расстояния определяется на основе одного атрибута или малого числа атрибутов. В более сложных случаях больших систем, для которых требуются сложные методы расчета меры расстояния, основанные на использовании набора атрибутов или на комбинации более простых мер расстояний, требуются и более сложные методы организации и индексации данных.

В качестве примера таких методов можно привести метод, использующий свойство неравенства треугольников (всякая сторона треугольника

больше или равна разности двух других сторон). В соответствии с ним можно записать неравенство для мер расстояний:

$$d(I, Q) \geq |d(I, K) - d(Q, K)|$$

для трех изображений:

- Q – изображение из запроса;
- I – изображение из базы данных;
- K – ключевое изображение.

В данном случае ключевое изображение позволяет найти нижнюю границу расстояния между изображениями из запроса и изображением из базы данных. Для этого из базы данных отбирают некоторый набор ключевых (эталонных) изображений K_1, K_2, \dots, K_m , вычисляют для них меру расстояния по отношению к изображению Q из запроса

$$d(Q, K_1), d(Q, K_2), \dots, d(Q, K_m).$$

Устанавливая некоторую границу T для меры расстояния между изображениями из базы данных и изображением из запроса, можно выполнить отбор изображений, удовлетворяющих требованиям пользователя:

$$|d(I, K_i) - d(Q, K_i)| \leq T.$$

Таким образом, выполняется уменьшение объема изображений из базы данных, необходимых для просмотра за счет исключения тех изображений, которые имеют недопустимую меру расстояния.

Для реализации данного подхода используется специальная структура данных, называемая треугольным деревом, которая позволяет при обработке запроса исключить из рассмотрения большое число изображений, для которых не производится сравнение с запросом.

На базе этого варианта разработан метод работы с мерами расстояний, которые определяются динамически в виде линейной или булевой комбинации базовых мер расстояний, в зависимости от условий задачи.



8. Обнаружение движения на изображениях

8.1. Задача обнаружения движения

Задача выявления движения и определения её параметров существует во многих сферах деятельности человека: охранные системы, системы контроля дорожного движения, геоинформационные системы, производственные конвейеры, робототехника и многие другие. Как правило, выявление движения на изображениях инициирует процедуру расчета характеристик движения, таких как форма движущегося объекта, направление и скорость движения, изменение взаимного положения движущихся объектов.

Факт наличия движения устанавливается на основе анализа изображений сцены, полученных через определенные промежутки времени. При этом наблюдаются изменения яркости и цвета в отдельных частях изображения. Движения на изображениях могут являться следствием как движения некоторых объектов сцены, так и движением самого наблюдателя (камеры) относительно неподвижного фона. В общем случае могут наблюдаться различные комбинации элементов динамики: камеры, объекты, фон:

- неподвижная камера, постоянный фон, один движущийся объект;
- то же с несколькими движущимися объектами;
- движущаяся камера, постоянный фон, один или несколько движущихся объектов.

Иногда для наблюдения движения используют несколько камер. В этом случае появляется возможность изучения трехмерных объектов и траектории их движения в пространстве. Съемка движущейся камерой дает значительно больше информации об окружающей среде, лучшей проработке глубины сцены и её составляющих объектов.

При обнаружении движения обычно решают задачи вычитания изображений, вычисление векторов перемещения, вычисление траекторий перемещения объектов и др.

8.2. Вычитание изображений

Вычитание изображений используют для обнаружения движущихся объектов на постоянном фоне. При этом два кадра будут отличаться только пикселями перемещения динамического объекта. Эти пиксели будут отличны от нуля, остальные – нулевые. Естественно, это наблюда-

ется в идеальном случае, без учета шумов, погрешностей съемки и т. д. В реальной ситуации логично вводить некоторый порог, превышение которого будет классифицироваться как наличие движения. Область, относящаяся к объекту движения, обычно обозначают прямоугольником для удобства дальнейшей обработки.

Алгоритм обработки изображения для выявления движущихся объектов можно построить следующим образом.

Введем обозначения:

- входное полутоновое изображение для момента времени $t - I_t(r, c)$;
- входное полутоновое изображение для момента времени $(t+\Delta) - I_{t+\Delta}(r, c)$;
- выходное изображение $- I_{out}(r, c)$;
- значение порога яркости μ .

Определим выходное изображение:

$$I_{out}(r, c) = \begin{cases} 1, & \text{если } |I_{t+\Delta}(r, c) - I_t(r, c)| > \mu; \\ 0, & \text{в противном случае.} \end{cases}$$

Далее на выходном изображении определяются все связанные компоненты, удаляются шумы и мелкие объекты, соизмеримые с шумами. Для близко расположенных связанных компонент выполняется замыкание с использованием структурирующего элемента в виде небольшого диска. В завершение вычисляются координаты описывающих прямоугольников, содержащих связанные компоненты. Результатом работы алгоритма будут выходное изображение $I_{out}(r, c)$ и массив координат описывающих прямоугольников.

8.3. Вычисление векторов перемещения

Для вычисления векторов перемещения объектов на изображении нужно установить одноименные точки на двух изображениях. В процедурах вычисления векторов перемещения используются некоторые термины, определения которых приводятся ниже.

Поле движения называется двумерный массив двумерных векторов, которые определяют перемещение точек трехмерной сцены. Векторы перемещения на изображении описывают смещения образов движущихся трехмерных точек. Начало этого вектора соответствует начальному положению трехмерной точки в момент времени t , а конец – положению этой точки в момент времени $(t + \Delta)$. Каждый такой вектор представляет собой мгновенную скорость точки в момент времени t .

Фокус расширения представляет собой точку на изображении, из которой расходятся все вектора поля движения. Точка на изображении, к которой направлены все вектора поля движения, называется **фокусом схода**. Точки фокуса расширения и фокуса схода можно также представить как точку, к которой движется камера, или точку, от которой удаляется камера соответственно.

Вычисление векторов перемещения основывается на предположении, что значения интенсивностей пикселей изображения и некоторой её окрестности в течение временного интервала (t_1, t_2) остается постоянной, или постоянна разность интенсивностей на образах краёв объектов.

Поток изображения представляет собой поле движения, вычисленное в предположении о постоянстве интенсивности изображений в окрестностях соответствующих точек.

Как уже говорилось, вектора движения можно вычислить по одноименным точкам двух изображений, найденных для моментов времени t и $(t+\Delta)$. Поиск соответствующих точек может зависеть от конкретной задачи. Нередко в качестве таких точек выбирают угловые точки соответствующих областей, или центр тяжести этих областей. Поиск угловых точек может производиться с помощью масок. Например, такими масками могут служить детектор краёв Кирша, или маски из базиса операторов Фрея – Чена [8], или признакового оператора. Для угловых точек этот оператор вычисляет изменение яркости по вертикальным, горизонтальным и диагональным направлениям соседей для пикселя P . Если наименьшая разность этих изменений интенсивности превышает установленный порог, то пиксель P считается признаковой точкой.

Определение признаковых точек может выполняться и другими методами. Например, аналогичный подход вместо вычисления разностей интенсивности опирается на вычисление текстурных разностей.

После обнаружения множества точек интереса $\{P_j\}$ на изображении I_1 для момента времени t , необходимо найти соответствующие точки на изображении I_2 для момента времени $(t+\Delta)$. Это можно сделать описанным выше способом или используя корреляционные методы найти точки интереса на изображении I_2 . Для этого выполняется исследование небольшой области для очередной точки интереса изображения I_1 с целью поиска подходящей точки на изображении I_2 с максимальной корреляцией. Найденные соответствующие точки интереса для двух моментов времени позволяют определить вектор перемещения.

8.4. Вычисление траекторий движущихся точек

Вычисление векторов перемещений дает представление о мгновенной скорости и направлении перемещений интересующей точки. Нередко в различных случаях ставится задача наблюдения за перемещением отдельной точки или объекта. Таким образом, ставится задача построения траектории движения. Наблюдение проводится на каком-то интервале времени через заданные интервалы отчета.

В простейшем случае, когда объект слежения обладает уникальной текстурой, используют метод вычисления нормированной кросс-корреляции. Сам объект может отслеживаться за счет его уникальности по текстуре или интенсивности.

В более общих случаях, когда объект не обладает характеристикой, присущей только ему, требуются иные подходы в регистрации движения интересующего пользователя объекта. В этих случаях используют какие-то другие особенности объекта, например особенности движения, размеры объекта, его структуру и другие. Кроме того, для упрощения расчетов могут накладываться определенные ограничения в наблюдении движущихся объектов. Например, известен такой набор предположений для движения физических объектов в трехмерном пространстве [1].

1. Движущийся объект в течение наблюдаемого времени перемещается непрерывно.
2. Скорость движущегося объекта с течением времени изменяется непрерывно и по величине, и по направлению.
3. В каждый момент времени объект может находиться только в одной области пространства.
4. Два различных объекта не могут находиться одновременно в одной области пространства.
5. Последнее допущение в общем случае может не выполняться при наблюдении объектов, которые в определенный момент времени могут перекрывать один другого. Это может потребовать дополнительных решений.

Построение траектории движения точки с учетом указанных выше предположений выполняется нахождением точек интереса на последовательности кадров и соединением их множеством гладких траекторий. Здесь траекторией объекта i называется последовательность точек $T_i = (p_{i,1}, p_{i,2}, \dots, p_{i,n})$, в которых наблюдается объект i в моменты времени $t = 1, 2, \dots, n$.

Для любых двух точек траектории можно определить вектор разности:

$$V_{i,t} = p_{i,t+1} - p_{i,t}.$$

Введем понятие гладкости в точке. Гладкость направления определяется скалярным произведением входящего и выходящего векторов в этой точке. Гладкость скорости определяется на основе вычисления геометрического и арифметического средних значений длин входного и выходного векторов по формуле

$$S_{i,t} = w \left(\frac{V_{i,t-1} \circ V_{i,t}}{|V_{i,t-1}| |V_{i,t}|} \right) + (1-w) \left(\frac{2\sqrt{|V_{i,t-1}| |V_{i,t}|}}{|V_{i,t-1}| + |V_{i,t}|} \right).$$

Весовой коэффициент отвечает условию $0 \leq w \leq 1$. Несложно показать, что гладкость скорости тоже отвечает этому условию $0 \leq S_{i,t} \leq 1$. Для прямолинейной траектории вектора разности для каждой точки равны между собой, а гладкость скорости $S_{i,t} = 1.0$, что соответствует максимальной гладкости в точке.

Если рассмотреть задачу построения траекторий для m уникальных точек (объектов) на интервале времени наблюдения в n кадров, то можно подсчитать их суммарную гладкость:

$$T_{sum} = \sum_{i=1}^m \sum_{t=2}^{n-1} S_{i,t}.$$

Алгоритм построения траекторий заключается в подборе точек на очередном кадре такого, что дает максимальное значение гладкости. На каждом шаге для траектории случайным образом назначают уникальные точки. Затем путем обмена точек проводят подсчет суммарной гладкости для каждого варианта и выбирают тот расклад, который дает максимальную суммарную гладкость. Такой алгоритм не гарантирует оптимальный результат, однако он достигает некоторого оптимума решения задачи. Объем вычислений можно оценить из расчета того, что на каждом кадре возможно выполнение $m(m-1)/2$ (число сочетаний из m по 2) операций обмена точек. Число таких кадров равно $(n-2)$. Поэтому общий объем операций обмена составляет величину

$$K = \frac{m(m-1)(n-2)}{2}.$$

Суммарная гладкость при этом не будет превышать величины $m(n-2)$. В реализации алгоритма можно предусмотреть различные допущения и упрощения, которые снижают объем вычислений.

Снижения объема вычислений можно добиться и за счет введения дополнительной информации о возможном поведении объекта. Например, движение живых существ имеет ограничение по скорости, по характеру движения составных частей тела. Для движения автомобилей можно учитывать массу, импульс движения, цвет автомашины и т. д. Дополнительная информация позволяет делать предсказание следующей точки траектории или более надежный выбор очередной точки.

8.5. Виды изменений условий съемки

Для создания цифровых библиотек видеоданных, адаптированных к задачам поиска, важной задачей является сегментирование и хранение видеопоследовательностей, которое облегчает произвольный доступ к ним. Ниже рассматриваются некоторые понятия и приемы, характерные для изменения сцен и условий съемки и которые можно использовать при сегментировании видеофайлов.

Смена сцены – характерная особенность для любых съемок, при которой существенно изменяется содержание сцены: фона, действующих объектов, динамики объектов. Переход от одной сцены к другой, как правило, занимает интервал 10–50 кадров с возможными различными эффектами перехода.

Смена монтажного кадра используется при съемке одной и той же сцены за счет переключения между несколькими камерами, снимающими разные части сцены или участвующих субъектов. Переключение производится достаточно быстро и без эффектов перехода.

Панорамирование камеры ставит своей задачей дать общую панораму сцены, на которой разворачиваются события. При этом камера движется по горизонтали в одном или другом направлении, а элементы сцены движутся в направлении, противоположном движению камеры.

Масштабирование применяется для показа деталей сцены или укрупнения общего плана сцены. Выполняется это изменением фокусного расстояния в течение некоторого времени.

Эффекты смены кадров применяются для перехода от одного ключевого кадра к другому, полученных от разных источников изображения. Набор таких эффектов очень широк. Можно привести примеры некоторых из них.

Плавное исчезновение кадра – это переход от некоторого изображения к пустому кадру (черному или белому). Противоположный эффект –

плавное появление кадра – переход от пустого кадра к некоторому изображению.

Замена одного изображения другим. Эта процедура может выполнять попиксельную замену одного изображения другим в случайном порядке. Замена может выполняться и блоками заданного размера.

Замена изображений может производиться вытеснением одного изображения другим путем перемещения границы их раздела в различных направлениях – по вертикали, горизонтали, диагонали. Новое изображение может, например, появляться в любой точке предыдущего изображения путем увеличения окна, в котором появляется новое изображение.



8.6. Сегментация видеопоследовательностей

Задача сегментации видеопоследовательностей заключается в разбиении последовательности видеок кадров на фрагменты, представляющие собой самостоятельные сцены или сюжеты. Сегментация может выполняться выявлением изменения качественного содержания сцен, наличием переходов между ними, изменением условий съемки, изменением количественных оценок двух кадров видеопоследовательности.

Количественные оценки двух кадров видеопоследовательности I_t и $I_{t+\Delta}$ можно определить вычислением средней разности между соответствующими пикселями двух изображений. Временной интервал Δ может занимать один или несколько кадров в зависимости от поставленной задачи и характера видеопоследовательности. Количественная оценка d различий двух кадров может быть рассчитана в соответствии со следующим выражением:

$$d(I_t, I_{t+\Delta}) = \frac{\sum_{r=0}^{MaxX-1} \sum_{c=0}^{MaxY-1} |I_t(r,c) - I_{t+\Delta}(r,c)|}{MaxX \times MaxY}$$

Данный расчет может дать неудовлетворительные результаты, так как даже незначительная смена кадров может приводить к существенным различиям в оценке. Поэтому могут использоваться и другие методы, приводящие к увеличению объема вычислений с одновременным повышением качества сегментации.

Другой метод основан на разбиении изображения на крупные блоки и установлением различий между соответствующими блоками двух изображений. Различия устанавливаются по значениям яркости с расчетами математического ожидания и дисперсии. При этом устанавливается некоторый порог по числу блоков совпадения. Если это число меньше порога,

то считается, что имеются в наличии одинаковые сцены или фрагменты видеопоследовательностей.

Еще один метод основан на вычислении разности гистограмм. На практике считается, что достаточно работать с 64-разрядной гистограммой. Для этого берут по два старших разряда от каждой компоненты цветности и получают 6-разрядное число, которое дает соответствующую 64-разрядную гистограмму. Во многих случаях данный подход обладает меньшим объемом вычислений и хорошим качеством сегментации. Однако в худшую сторону на результат может повлиять разное пространственное распределение значений яркости.



9. Сопоставление изображений в двумерном пространстве

9.1. Основные понятия

Сравнение изображений может выполняться в различных вариантах: сравнение двух изображений, сопоставление изображения и карты, наконец, сравнение изображений и моделей объектов. В каждом из описанных случаев имеются свои особенности и методы их решения. Области использования этих операций – распознавание образов, построение трехмерных сцен по двумерным изображениям, картография, медицина и другие. Сопоставление изображений, карт или моделей объектов заключается в сопоставлении точек одного изображения точкам другого. Математически указанное отображение модели M и объекта I может быть представлено в следующем виде:

$$M(x, y) \cong I(g(x, y), h(x, y));$$
$$I(r, c) \cong M(g^{-1}(r, c), h^{-1}(r, c)).$$

Отображение одного двумерного координатного пространства на другое, в общем случае, называется двумерным преобразованием. Отображающие функции g и h задают соответствие между модельными точками и точками изображения. Эти функции позволяют по любой точке модели вычислить координаты соответствующей точки изображения. Отображающие функции являются обратимыми, поэтому с помощью обратных функций можно выполнять преобразование в двух направлениях.

Преобразование точек двух изображений, при котором точки двух изображений получают одинаковые координаты, называется совмещением изображений. При этом предполагается, что изображения содержат одну и ту же сцену.

В зависимости от вида отображающих функций g и h получают разные преобразования – линейные, нелинейные, непрерывные, дискретные и другие. В графике чаще всего используют аффинные преобразования, а точки на плоскости и в пространстве представляются однородными координатами. Аффинное преобразование характеризуется тем, что оно переводит параллельные линии в параллельные, пересекающиеся прямые – в пересекающиеся.

Аффинные преобразования могут выполнять преобразования точки следующего вида – перенос, масштабирование, поворот, сдвиг, деформация и т. д.

В общем случае установить параметры композиции преобразований, которая переводит координаты контрольных точек модели в координаты точек изображения, можно путем решения системы уравнений. Например, композиция преобразований содержит операции переноса точек, масштабирования и поворота. Тогда можно записать уравнение преобразования точки. Матрицы отдельных преобразований имеют значения:

$$\text{для переноса} \quad T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ dx & dy & 1 \end{bmatrix};$$

$$\text{для масштабирования} \quad S = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

$$\text{для поворота} \quad R = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Матрица композиции преобразований переноса, масштабирования и поворота определяется перемножением указанных матриц отдельных преобразований. Она позволяет записать соотношения для соответствующих координат точек модели и изображения:

$$\begin{aligned} x_m &= x_i s \cos\alpha - y_i s \sin\alpha + dx; \\ y_m &= x_i s \sin\alpha + y_i s \cos\alpha + dy. \end{aligned}$$

Если расчеты выполнять для отрезка прямой, то можно записать систему четырех уравнений для двух концевых точек отрезка. При этом угол наклона отрезка в каждой системе координат определяется выражением

$$\beta = \operatorname{arctg} \frac{y_2 - y_1}{x_2 - x_1}.$$

Тогда угол α между отрезком модели и отрезком изображения будет

$$\alpha = \beta_m - \beta_i.$$

Зная угол между отрезками, можно решить указанную выше систему уравнений и получить значения s , dx , dy .

9.2. Распознавание двумерных объектов с помощью аффинных преобразований

Метод базовых локальных признаков. При распознавании этим методом учитываются локальные характерные признаки объектов и их взаимное расположение. Для каждого распознаваемого объекта составляется модель, которая содержит множество базовых признаков. Базовые

признаки являются главными характерными признаками объекта и должны легко обнаруживаться в процессе распознавания.

Кроме того, для каждого объекта определяется множество близкорасположенных характерных признаков, которые служат для подтверждения корректного обнаружения базового признака и определения местоположения и ориентации объекта.

Алгоритм сопоставления работает следующим образом. Сначала выделяются характерные признаки для одного или нескольких объектов изображения. В случае обнаружения базового признака одной из моделей алгоритм ищет в окрестности базового признака близкорасположенные признаки изображения. Если на изображении удастся обнаружить признаки, похожие на модельные близкорасположенные признаки, то между признаками изображения и модели устанавливается соответствие. В соответствии с этим высказывается гипотеза о присутствии на изображении известного объекта. После этого требуется подтверждение данной гипотезы путем выполнения процедуры верификации.

Цель проведения верификации заключается в проверке достаточности оснований для того факта, что предполагаемый объект присутствует в изображаемой сцене. С учетом множества установленных соответствующих характерных признаков вычисляются параметры аффинного преобразования, которое переводит модельные точки в точки изображения. С помощью полученных отображающих функций все прямолинейные сегменты границы модели объекта переводятся в координаты концов отрезков на изображении. Преобразованные отрезки на изображении примерно должны совпадать с отрезками модели. При этом следует учитывать влияние шумов, погрешностей вычислений и сопоставлений признаков. Поэтому поиск подходящих отрезков на изображении следует искать в некоторой прямоугольной области вокруг каждого преобразованного модельного отрезка. Если удастся найти существенные признаки совпадения, то модельный отрезок помечается как подтвержденный, а если окажется, что таких подтвержденных модельных отрезков будет в достаточном количестве, то объект считается обнаруженным. Параметры отображающих функций определяют расположение объекта на изображении.

Кластеризация по расположению. Выше было показано, как устанавливаются параметры отображающих функций для соответствующих контрольных точек модели и изображения. Однако при таком подходе самым важным становится качество выбора соответствующих контрольных точек. Поэтому в другом методе кластеризации по расположению задача выбора таких точек решается вычислением аффинных преобразо-

ваний для всех пар точек модели и изображения. После этого все множество полученных параметров аффинных преобразований выполняет поиск кластера, содержащего наиболее близкие значения. При этом выполняется анализ и по сопоставлению признаков точек этого кластера. Если их достаточно много, то считается, что сопоставление точек модели и изображения выполнено.

Под расположением здесь понимаются параметры местоположения и ориентации объекта, определяемые преобразованием, совмещающим модель с образом объекта на изображении. Данный подход характеризуется значительным объемом вычислений. Поэтому можно использовать дополнительные характеристики и признаки, которые дают возможность уменьшить число вычислительных операций. Например, если говорить об аэрофотосъемках, то при сопоставлении модели и изображении удобно использовать характерные особенности местности и рисунки дорог, железнодорожных путей, разбиения на поля. Эти признаки точки и сегменты позволяют отбрасывать многие парасочетания объектов или их элементов, как не удовлетворяющие задаче поиска. Поэтому оценка объема вычислений метода по расположению $O(n^2)$, где n – число сопоставимых наборов, может быть существенно снижена.

Геометрическое хэширование. Рассмотренные методы базовых локальных признаков и алгоритм кластеризации по расположению предназначены для простых случаев сопоставления одной модели и изображения. При наличии нескольких моделей возникает необходимость сопоставления всех возможных пар сочетаний моделей и изображения, что приводит к росту вычислительных затрат.

Для таких случаев разработан метод геометрического хэширования, который требует большой подготовительной обработки данных, но с последующим быстрым распознаванием и определением параметров расположения объекта.

Задача ставится следующим образом. Имеется большой набор моделей и некоторый объект с набором выделенных характерных признаков. Считается, что этот объект может быть найден среди моделей путем аффинного преобразования одной из них. Поэтому решение задачи поиска – найти параметры аффинного преобразования, совмещающего модель с объектом на изображении.

Пусть задана модель M в виде упорядоченного множества признаков точек. Если взять из этого множества три неколлинеарных точки $A = \{a_0, a_1, a_2\}$, то они будут определять аффинный базис модели M , а следовательно, и систему координат (рис. 21).

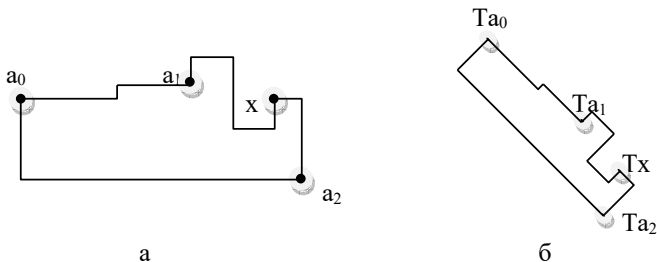


Рис. 21. Аффинное преобразование объекта:
а – исходный объект; б – преобразованный объект.

С помощью выбранной системы координат любую точку $x \in M$ можно представить с помощью аффинных координат (α, β) :

$$x = \alpha(a_1 - a_0) + \beta(a_2 - a_0) + a_0.$$

Применяя аффинное преобразование T , получим аффинные координаты точки x преобразованного объекта:

$$Tx = \alpha(Ta_1 - Ta_0) + \beta(Ta_2 - Ta_0) + Ta_0.$$

На этапе предварительной обработки формируется хэш-таблица со всеми моделями базы данных. В качестве индекса хэш-таблицы выступают аффинные координаты (α, β) , и каждая строка содержит пару модель-базис (M, A) .

После составления хэш-таблицы выполняется операция распознавания объектов. Сначала создается аккумуляторный массив V , проиндексированный парами модель-базис, с нулевыми значениями для каждой пары (M, A) . В процессе распознавания они получают значения по числу голосов, поддерживающих гипотезу о том, что существует преобразование T , отображающее пару (M, A) на изображение.

На заключительном этапе выполняется расчет параметров преобразований для тех пар модель-базис, для которых было получено достаточно большое число голосов.

На этапе предварительной обработки оценка сложности расчетов для s моделей, каждая из которых содержит n точек, составляет $O(sn^4)$, т. е. ведется обработка $O(sn^3)$ троек значений и $O(sn)$ других точек. Конкретная сложность расчетов зависит от ряда условий. Это удачное нахождение признаков точек, обнаружения ложных и лишних признаков точек. При самом удачном раскладе сложность может составить $O(sn)$ операций, а в худшем случае – $O(sn^4)$. Диапазон количества операций достаточно широк, и поэтому очень важно сокращение вычислений за счет дополнительных приемов повышения точности расчетов на всех этапах

обработки и распознавания. При этом следует учитывать, что ошибки могут возникать за счет неточного определения координат признаков точек, пропуска признаков точек или обнаружения лишних. Задача распознавания может существенно осложняться за счет перекрытия нескольких объектов и ошибок при аффинных преобразованиях для базисных троек.

9.3. Распознавание двумерных объектов с использованием реляционных моделей

При распознавании двумерных объектов на основе реляционных моделей вводятся следующие понятия.

Часть – это структурный компонент сцены (объект, сегментированная область, контур, отверстие и т. п.).

Метка – символ, идентифицирующий некоторую часть.

Назначение – отображение частей на метки. В простейшем случае, например, назначение – это нумерация объектов сцены.

Отношение – абстрактное математическое понятие. Оно устанавливается между объектами сцены и может храниться для модельных объектов.

Согласованной маркировкой f называется такое назначение меток частям, при котором выполняется условие:

$$\text{если } (p_i, p_j) \in R_P, \text{ то } (f(p_i), f(p_j)) \in R_L,$$

где p_i, p_j – части;

L – множество меток частей;

R_P, R_L – отношения на множествах P и L соответственно.

Проиллюстрируем сказанное на примере двух изображений. Пусть даны множества отрезков P одного изображения и множество отрезков L другого изображения (рис. 22).

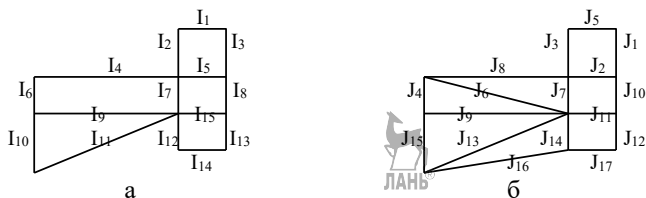


Рис. 22. Согласованная маркировка двух изображений:
а – изображение I; б – изображение J.

В таблице 7 приводится согласованная маркировка двух изображений.



Таблица 7

Сегменты изображения I_k	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}
Сегменты $f(I_k)$ изображения J	J_5	J_3	J_1	J_8	J_2	J_4	J_7	J_{10}	J_9	J_{15}	J_{13}	J_{14}	J_{12}	J_{17}	J_{11}

Кроме указанных в таблице 7 сегментов, изображение J содержит сегменты, у которых нет аналогов на изображении I – это J_6 и J_{16} .

В данном примере дается согласованная маркировка, которую на практике требуется установить. Причем в общем случае сегменты могут соответствовать друг другу приблизительно из-за шумов, неточности расчетов и других причин. Поэтому нужен какой-то допустимый диапазон отклонений в размерах и расстояниях.

Задача определения согласованной маркировки может выполняться построением **дерева интерпретации**, которое представляет все возможные варианты назначения меток частям объектов. При построении дерева интерпретации учитываются ограничивающие отношения, что упрощает построение дерева, и уменьшает его размеры.

Задачей распознавания в данном случае является установление согласованной маркировки f . Дерево интерпретации строится с помощью рекурсивной обработки с возвратом. При этом формируется путь просмотром дерева «в глубину». До окончания процедуры распознавания текущий параметр f содержит неполное согласованное значение. Первоначальное его значение является нулевым. На каждом шаге обработки, если новая маркировка части согласуется с неполным назначением, выполняется попытка назначения непомеченной части другой метки. В противном случае, при несогласованности с неполным назначением, производится возврат алгоритма и выбор альтернативного варианта.

Другой подход определения согласованной маркировки носит название **дискретной релаксации**. Он отличается некоторым упрощением учета имеющихся ограничений. Упрощение заключается в том, что учитываются только локальные ограничения. Это приводит к тому, что снижается объем вычислений и появляется возможность параллельной обработки дерева интерпретации из-за независимости обработки на одном пути дерева интерпретации (в пределах локальности ограничений).

В соответствии с алгоритмом каждой части из множества обнаруженных характерных признаков изображения согласно ее типу назначаются все множество допустимых меток. После этого выполняется провер-

ка допустимости всех отношений. Обнаруженные несогласованные метки удаляются из множества допустимых меток. При этом, если такое удаление производилось, выполняется еще один шаг обработки. Иначе процесс фильтрации прекращается. Обработку множества меток для разных частей можно выполнять параллельно.

Работу алгоритма проиллюстрируем на следующем примере. На рисунке 23 изображена модель объекта, у которого имеются отверстия. Расположение отверстий и расстояния между их центрами приведены в таблице 8.

Таблица 8

Точка	Коорд. центра	Расстояния между центрами отверстий					
		A	B	C	D	E	F
A	(5, 11)	0	3,2	7,3	5,4	7,6	7,8
B	(8, 12)	3,2	0	9,4	6,1	5,7	6,7
C	(3, 4)	7,3	9,4	0	4,5	9,8	8,2
D	(7, 6)	5,4	6,1	4,5	0	5,4	4,0
E	(12, 8)	7,6	5,7	9,8	5,4	0	2,2
F	(11, 6)	7,8	6,7	8,2	4,0	2,2	0

В результате выделения характерных признаков на изображении были обнаружены только центры трех отверстий объекта. Пространственные преобразования объекта для сопоставления с характерными признаками изображения допускают переносы, повороты и масштабирования.

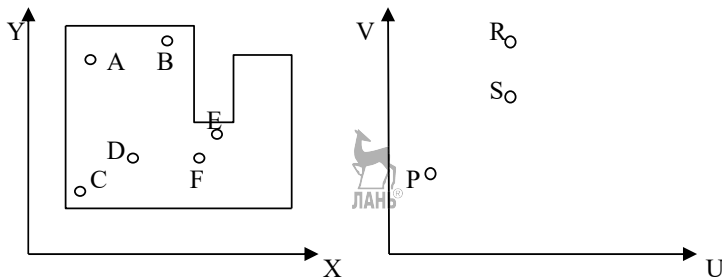


Рис. 23. Модель объекта (слева) и обнаруженные признаки на изображении (справа)

Для данного примера считается, что коэффициент масштабирования известен, и поэтому к изображению применено масштабирование, после которого получено изображение, показанное на рисунке 23 (справа).

Аналогичная таблица 9 составлена для обнаруженных точек изображения.

Таблица 9

Точка	Координаты центра	Расстояния между центрами отверстий		
		P	R	S
P	(2, 5)	0	7,8	5,4
R	(7, 11)	7,8	0	4,0
S	(7, 7)	5,4	4,0	0

На основании двух таблиц выполним фильтрацию. Для этого каждому из трех отверстий изображения назначаются шесть возможных меток модели. В результате анализа допустимости этих назначений удаляем те, которые не удовлетворяют таким назначениям по каким-то причинам. Результаты первого прохода маркировки методом дискретной релаксации приведены в таблице 10.

Таблица 10

	A	B	C	D	E	F
P	$d(P, R) = 7,8;$ $d(P, S) = 5,4$	нет $d = 7,8;$ $d = 5,4$	нет $d = 7,8;$ $d = 5,4$	нет $d = 7,8$	нет $d = 7,8$	нет $d = 5,4$
R	нет $d = 4,0$	нет $d = 7,8;$ $d = 4,0$	нет $d = 7,8;$ $d = 4,0$	нет $d = 7,8$	нет $d = 7,8;$ $d = 4,0$	$d(R, P) = 7,8;$ $d(R, S) = 4,0$
S	нет $d = 4,0$	нет $d = 5,4;$ $d = 4,0$	нет $d = 5,4;$ $d = 4,0$	$d(S, P) = 5,4;$ $d(S, R) = 4,0;$ $d(S, E) = 5,4$	нет $d = 4,0$	нет $d = 5,4$

В таблице 10 указаны не подходящие по причине отсутствия допустимых расстояний между точками назначения меток. В случае допустимости назначений в клетках таблицы указаны варианты назначений. Так, в ячейке на пересечении строки P и столбца A расстояние $d(P, R) = 7,8$ удовлетворяет назначению $F \in L(R)$, а расстояние $d(P, S) = 5,4$ – назначению $D \in L(S)$. Этот же результат подтверждается и ячейками таблицы на пересечении строки R и столбца F, а также строки S и столбца D. В последней ячейке (S/D) расстояние $d(S, E) = 5,4$ не согласуется ни с какими назначениями, поэтому оно фильтруется как не удовлетворяющее условиям поиска.

Ввиду простоты примера фильтрация заканчивается уже после первого прохода. Приведенный пример рассматривался с точки зрения идеального случая, когда расстояния и другие параметры модели и изображе-

ний в точности совпадают. В реальных условиях следует учитывать, что соответствующие значения могут иметь некоторую величину отклонения. В этом случае задача усложняется, и могут появиться дополнительные варианты назначений, требующих анализа и более широкого просмотра различных назначений.

Сопоставление по реляционному расстоянию. Метод сопоставления по реляционному расстоянию применяется в случаях, когда особенно важно получить оптимальное решение. Решение задачи по этому методу строится поиском наилучшего отображения f на основе критерия выполнения большинства отношений или по критерию минимизации количества пустых меток [9, 10].

С этой целью строится реляционное описание модели, в котором содержатся части описываемой сущности и отношения, указывающие на наличие различных взаимосвязей между частями. Реляционное описание представляет структуру данных, с помощью которой можно выполнять сопоставление моделей различных двумерных и трехмерных объектов, областей изображения и других.

Реляционное описание позволяет найти сходства и различия сопоставляемых моделей и изображений, которые оцениваются общей ошибкой преобразования в виде суммы структурных ошибок по каждому виду отношений. Выбор наиболее подходящего отображения f выполняется по минимальному реляционному расстоянию между двумя реляционными описаниями, которое представляет собой минимальную полную ошибку некоторого отображения f . Это минимальное реляционное расстояние соответствует наилучшему отображению сопоставляемых объектов.

Структурная индексация. Для повышения надежности и скорости поиска нужных изображений нередко используют реляционную индексацию. Она состоит в том, что выполняется поиск всех изображений, которые содержат выбранный структурный элемент. Эта процедура поиска может повторяться многократно для разных структурных элементов. После этого можно ставить задачу выбора из базы данных изображений, которые содержат указанные структурные элементы или их часть. Задача решается голосованием. Те изображения, которые набирают наибольшее число голосов, будут обладать наибольшим сходством с моделью, определенной заданием поиска. В зависимости от постановки задачи можно отбирать изображения, содержащие указанный набор структурных элементов или часть из них в определенном диапазоне.

10. Распознавание объектов на изображениях

10.1. Определение соответствия точек двумерного изображения и точек трехмерной сцены

В задачах компьютерного зрения нередко возникает необходимость установления связи между структурами на двумерных изображениях и соответствующими структурами трехмерного пространства. Иначе говоря, необходимо построить картину трехмерной сцены по одному или нескольким двумерным изображениям.

Такая постановка задачи характерна для систем распознавания и контроля трехмерных объектов, информационных систем робототехники, автоматизированных производств, систем виртуальной реальности и других областях.

Для построения точки трехмерного пространства по изображениям, полученным от двух камер, необходимо знать параметры этих камер и установления соответствия точек трехмерного пространства и точек двух изображений. В качестве параметров камер служат координаты их положения и фокусные расстояния. Имея указанные исходные данные, можно провести лучи из камер к соответствующей точке пространства и выполнить несложные расчеты для определения трехмерных однородных координат точки $P = (X, Y, Z, W)$. Для выполнения указанных расчетов обычно используют четыре системы координат: система координат модели, мировая система координат, системы координат двух камер. Переход от одной системы координат к другой можно выполнить, если известны характеристики их взаимного расположения – перенос начала одной системы координат относительно другой и взаимные повороты осей координат. Операции преобразований точки в трехмерном пространстве переноса, поворота, масштабирования и их композиций [11] позволяют вычислять необходимые аффинные преобразования для вычисления координат точек пространства.

10.2. Распознавание объектов на основе моделей

Основным методом распознавания объектов на изображениях является использование моделей трехмерных объектов. Модели должны обла-

дать простой структурой для удобства расчетов и обработки, а также достаточно близко соответствовать распознаваемому объекту. Одним из самых распространенных видов моделей является каркасное представление трехмерного объекта, которое содержит в основном вершины и соединяющие их ребра. При этом вершины представляют собой точки трехмерного пространства. Такие модели удобны для представления трехмерных объектов, образованных плоскими гранями. Ввиду широкого распространения каркасных моделей многие графические аппаратные устройства имеют соответствующую поддержку для них.

Для описания объектов, имеющих гладкие кривые поверхности, используют модели в виде кривых второго порядка. Для этих целей разработаны кривые Безье, В-сплайны и другие аналитически описываемые поверхности.

В некоторых случаях возможно использование комбинаций разных видов моделей или моделей специального вида, адаптированных для конкретных объектов. Модели, отображающие пространственные характеристики объектов, нередко снабжаются дополнительной информацией в виде цвета, текстуры, отражающих свойств поверхностей и другие. С обработкой таких моделей успешно справляются современные графические процессорные устройства.

Задачи компьютерного зрения требуют такого представления объектов, которое позволяет успешно решать задачу распознавания, т.е. надежно устанавливать связь между описанием модели объекта и характерными признаками, содержащимися в изображении объекта. В зависимости от форматов используемых изображений (полутоновые цветные, дальнометрические) разрабатываются модели для задач конкретного типа. Ниже будут рассмотрены более подробно разновидности таких моделей.

10.3. Существующие разновидности трехмерных моделей

Трехмерные каркасные модели. Самый простой и хорошо обрабатываемый вид моделей в виде трехмерного каркаса. Такая модель состоит из плоских граней в виде различных многоугольников, ориентированных соответствующим образом в пространстве. Каждая грань содержит вершины трехмерной модели и соединяющие их ребра.

Если модель содержит многоугольники одного типа (треугольники, четырехугольники, пятиугольники и т. д.), то модель называется регулярной каркасной моделью. Наиболее распространенный тип каркасных

моделей – триангулированные каркасные модели. Такие модели удобны для обработки и позволяют добиваться нужной точности представления реальных объектов соответствующей аппроксимацией.

Модели типа «поверхность – ребро – вершина». Модели этого типа отличаются от каркасной модели тем, что ребра могут быть в виде кривых, а грани – в виде криволинейных поверхностей. Ребра в виде кривых в простейших случаях могут представляться окружностями или дугами, а криволинейные поверхности – в виде сфер, полусфер, цилиндров, конусов и т. д. В более сложных случаях элементы трехмерных объектов могут иметь более сложные представления, описываемые аналитически.

Модели на основе обобщенных цилиндров. Обобщенный цилиндр – это объемный элемент, заданный пространственной осью и функцией поперечного сечения для каждой точки оси. Пространственная ось может иметь разные формы, каждая из которых будет задавать свою модель в виде обобщенного цилиндра. В это разнообразие вносит свой вклад и функция сечения. В качестве простых примеров можно привести следующие модели.

Цилиндр. Пространственная ось – прямая линия, поперечное сечение – окружность постоянного радиуса.

Конус. Пространственная ось – прямая линия, поперечное сечение – окружность с радиусом в интервале от 0 в начале оси до некоторого максимального значения в конце оси.

Тор. Пространственная ось – окружность, поперечное сечение – окружность постоянного радиуса.

Параллелепипед. Пространственная ось – прямая линия, поперечное сечение – прямоугольник фиксированного размера.

Аналогично можно определить такие трехмерные объекты, как призма, пирамида (прямые и наклонные) и т. д. Можно представить с помощью обобщенных цилиндров и более сложные объекты (например, человека, животных) из комбинаций различных обобщенных цилиндров. При этом можно добиваться разной степени детализации представления реального объекта.

Модель объекта на основе обобщенных цилиндров содержит описание его составных частей, тоже представленных обобщенными цилиндрами. При этом даются параметры каждого такого обобщенного цилиндра, которые могут включать длину оси, средний размер поперечного сечения, отношения двух сечений, угол конусности и другие. Кроме того, указываются пространственные параметры всего объекта и связность со-

ставных частей. Пространственные отношения частей объекта могут включать координаты точек их соприкосновения, углы в точках связи, углы поворота обобщенных цилиндров.

Детализация объекта предполагает наличие разных уровней – от первого, грубого представления, до последнего с предусмотренной максимальной детализацией. Модели разных уровней отличаются степенью проработки каждой составной части модели объекта. Например, рука человека на первом уровне может состоять из одного обобщенного цилиндра. На следующем она может быть представлена в виде трех обобщенных цилиндров, изображающих предплечья и кисть. На последующем уровне можно детализировать кисть в виде 5 пальцев. Число уровней зависит от поставленной задачи и технических возможностей моделирования.

Если обратиться к двумерным изображениям, то обобщенные цилиндры дают проекции на изображениях в виде лент и эллипсов. Ленты возникают от проекции протяженной части цилиндра, а эллипсы – от поперечных сечений цилиндров. В идеале эти проекции могут иметь форму прямоугольника и эллипса. В общем же случае они приближаются к этим формам с разной степенью точности. При решении задачи распознавания объектов на изображении можно использовать характерные особенности моделирования в виде обобщенных цилиндров.

Модели на основе октантных деревьев. Октантное дерево представляет собой древовидную структуру данных, у которой каждый узел имеет восемь дочерних узлов. Узлы соответствуют некоторому кубическому объему трехмерного пространства. Метка, соответствующая узлу, может находиться в трех состояниях: заполнено, пусто, частично заполнено. Эти состояния определяются по факту принадлежности соответствующего куба объекту. Узлы, имеющие метки со значениями «полный» или «пустой», не имеют дочерних узлов. Узлы с частичным заполнением разбиваются на 8 дочерних узлов следующего уровня октантного дерева.

Такая структура трехмерного объекта описывается трехмерным массивом размера $2^n \times 2^n \times 2^n$, где n – некоторое целое число. Элементы этого массива называются вокселями (объемными элементами) и могут принимать значения 0 или 1, т. е. пустой или полный объемный элемент.

Приведем простой пример октантного дерева для объекта, представленного на рисунке 24.

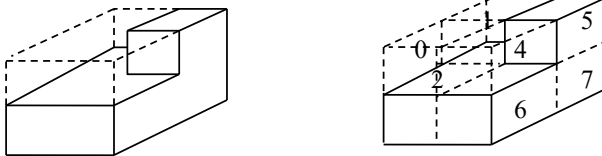


Рис. 24. Пример объекта для октантного дерева

Октантное дерево, представляющее структуру данного объекта, приведено ниже (рис. 25).

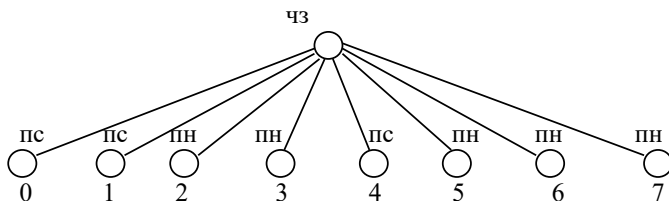


Рис. 25. Октантное дерево

Здесь корневой узел представляет исходный объем (на рис. 24 слева) в форме параллелепипеда, построенного на минимаксных координатах объекта. На дереве он помечен меткой «чз» – частично заполненный. Восемь дочерних узлов, пронумерованных 0–7, соответствуют восьми объемам деления объекта (на рис. 24 справа). Воксели 0, 1 и 4 являются пустыми (помечены метками «пс»), а воксели 2, 3, 5, 6, 7 являются полными (метки «пн»). Так как дочерние узлы не содержат метки «чз», то процесс построения дерева на этом заканчивается.

Модели на основе суперквадрик. Модели этого типа представляют собой семейство параметрических поверхностей, которое может давать высокую степень приближения к моделируемому объекту. Суперквадрика может быть представлена вектором $S(\alpha, \beta) = (x, y, z)$, компоненты которого определяются как функции углов α и β следующими соотношениями:

$$\begin{aligned} x &= a_1 \cos^p(\alpha) \cos^r(\beta), \\ y &= a_2 \cos^p(\alpha) \sin^r(\beta), \\ z &= a_3 \sin^p(\alpha). \end{aligned}$$

Здесь углы $-\frac{\pi}{2} \leq \alpha \leq \frac{\pi}{2}$ и $-\pi \leq \beta < \pi$. Параметры a_1, a_2, a_3 определяют размеры суперквадрики по осям x, y и z , а параметры p и r – «прямоугольность» в плоскостях широты и долготы.

С помощью суперквадрик можно описывать широкий набор простейших моделей, с помощью которых составлять более сложные объекты. Примерами таких простейших блоков могут быть сферы, цилиндры, параллелепипеды, эллипсоиды и другие, а также деформации из этих объектов. Например, при значении параметров $p = r = 1$ суперквадрика дает эллипсоид, а при равенстве параметров $a_1 = a_2 = a_3$ эллипсоид переходит в сферу. Наконец, при $p \ll 1$, а $r = 1$ поверхность будет походить на цилиндр.

Основное достоинство суперквадрик состоит в том, что они позволяют создавать деформируемые геометрические объекты операциями сужения и изгиба.

Операция линейного сужения вдоль оси z может быть получена следующими преобразованиями:

$$\begin{aligned} x^* &= \left(\frac{k_x}{a_3} z + 1 \right) x; \\ y^* &= \left(\frac{k_y}{a_3} z + 1 \right) y; \\ z^* &= z. \end{aligned}$$

Коэффициенты сужения k_x и k_y в плоскостях x и y по отношению к оси z могут изменяться в пределах $-1 \leq k_x, k_y \leq 1$.

Вторая операция деформации изгиба выполняется преобразованием

$$\begin{aligned} x^* &= x + \cos\theta(R - r); \\ y^* &= y + \sin\theta(R - r); \\ z^* &= \sin\gamma \left(\frac{1}{k} - r \right), \end{aligned}$$

где k – кривизна, а r – проекция x - и y -компонент на плоскость изгиба $z-r$:

$$r = \cos \left(\theta - \operatorname{tg}^{-1} \left(\frac{y}{x} \right) \right) \sqrt{x^2 + y^2}.$$

Преобразование R описывается выражением

$$R = k^{-1} \cos\gamma (k^{-1} - r),$$

а угол изгиба $\gamma = zk^{-1}$.

Хороший подбор параметров сужения и изгиба может дать качественное приближение модели к сложным реальным объектам.

10.4. Классы видимости трехмерных объектов

Рассмотренные выше модели, используемые для представления трехмерных объектов, связаны с характеристиками пространственных форм объектов. В задачах компьютерного зрения распознавание объектов на изображениях рассматривает также классы точек наблюдения объектов.

Задача ставится таким образом, что всевозможные точки наблюдения объекта разбиваются на множество классов видимости. Каждый класс содержит такие точки, которые дают проекции объекта на изображении с одинаковыми свойствами. Такими свойствами проекций могут быть одинаковые грани объекта, одинаковые прямолинейные отрезки, малое относительное расстояние между видами изображения и другие.

Если для объектов разбиение на классы видимости выполнено корректно, то решение задачи сопоставления объекта и изображения существенно облегчается за счет полученных ограничений принадлежности изображения к определенному классу видимости.

Пример классов видимости для треугольной призмы показан на рисунке 26.

Здесь даны 4 класса видимости для треугольной призмы. Например, можно отметить, что класс 1 содержит все точки пространства, которые лежат на прямой перпендикулярной плоскости проекции и проходящей через центр прямоугольника. Аналогично можно определить состав и других классов.

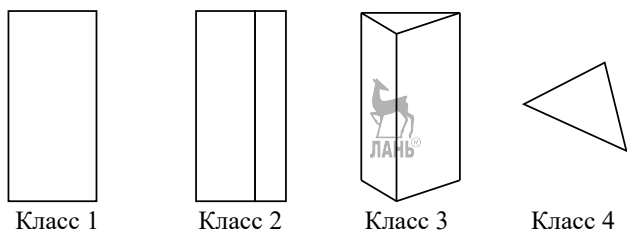


Рис. 26. Классы видимости для призмы

10.5. Физические и деформационные модели

Физические модели могут использоваться для моделирования внешнего вида и поведения физического объекта для распознавания их на изображении. С понятиями физической модели, как правило, связывают

и понятие деформационной модели, которая демонстрирует изменение модели во времени.

Модели в виде активных контуров. Такие модели используют для описания поведения деформируемых физических объектов под воздействием внешних и внутренних сил. Примерами моделей с активными контурами могут быть футбольный мяч, замкнутая резиновая лента вокруг фиксированных физических объектов, автомобильная камера, воздушный шарик, пружина и другие. Движущей силой деформации у таких объектов может быть давление сжатого воздуха, силы упругости, энергия взрыва и т. п. Как правило, деформация происходит в течение какого-то времени до уравнивания сил, воздействующих на объект. После этого объект принимает статический вид.

Для моделирования поведения активного контура определяется структура данных о форме, положении контура, силах, воздействующих на него. На рисунке 27 показана схема для динамики активного контура для двух случаев: в одном – движение вовнутрь под воздействием силы упругости (например, резиновая лента), в другом – движение наружу под воздействием силы, например, сжатого воздуха (мяч, автомобильная камера).

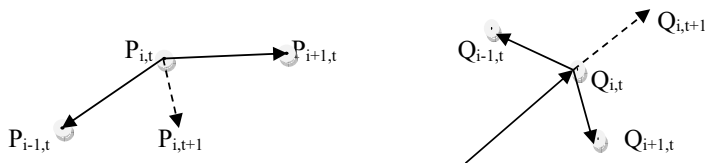


Рис. 27. Схема поведения активного контура

На рисунке 27 слева точка $P_{i,t}$ ускорено движется в положение $P_{i,t+1}$. Для простоты расчета считается, что каждая точка обладает единичной массой. Зная ускорение, можно вычислить скорость перемещения и новые координаты точки. Это движение продолжается до тех пор, пока не встретит жесткое ограничение (препятствие). Аналогичные расчеты можно выполнить и для случая, изображенного на рисунке 27 справа. Только здесь процесс движения заканчивается уравниванием сил, действующих на точку.

Оболочечные модели. Этот вид моделей используется для построения каркасных моделей объектов сложной конфигурации. Принцип создания модели состоит в следующем. Исходными данными моделируемого объекта являются трехмерные точки, принадлежащие поверхности

объекта в виде облака. Далее строится триангулированная сферическая модель и помещается внутрь облака исходных трехмерных точек. Затем каждая вершина сферы начинает перемещаться под действием силы растяжения в направлении нормали к поверхности сферы. Движение вершин происходит до момента достижения точек облака. Те вершины, которые не достигли точек облака, продолжают свое движение. Если треугольная грань в результате расширения становится большой, то она разбивается на более мелкие треугольники. Такой процесс расширения модели выполняется до тех пор, пока все вершины модели не достигнут трехмерных точек облака. Таким образом модель заполняет все облако и принимает вид каркасного триангулированного объекта. Направление нормали к поверхности модели можно определять усреднением нормалей к граням, которые сходятся в данной вершине.

Этот метод построения каркасной модели объекта с точки зрения вычислений проще, чем непосредственное формирование поверхности объекта по облачным точкам исходных данных.

10.6. Методы распознавания трехмерных объектов

10.6.1. Классификация задач распознавания

В общем виде задача распознавания является достаточно сложной и не имеющей одного универсального метода решения. Поэтому возникает необходимость разбиения задачи распознавания на ряд частных задач, каждая из которых имеет определенные ограничения, упрощающие решение задачи распознавания. Такое разбиение можно выполнить по различным критериям, которые рассматриваются ниже.

Область распознавания. Задачи распознавания, относящиеся к различным сферам приложения, могут сильно различаться по постановке задачи, исходным данным, точности решения и другим параметрам. Поэтому задачи распознавания могут относиться к техническим, психологическим, биологическим, изобразительным областям и т. д. Требования к решению задачи распознавания в каждой из этих областей будут существенно различаться между собой и, тем самым, инициировать различные методы распознавания.

Объекты распознавания. Основным принципом классификации объектов по их типу – деление на объекты искусственного и естественно-го происхождения. Объекты искусственного происхождения отличаются большей определенностью, большей регулярностью, что позволяет со-

ставлять подходящий набор параметров, хорошо описывающих объект. Распознавание таких объектов представляет собой относительно более простую задачу по сравнению с объектами естественного происхождения. Характеристики объектов естественного происхождения отличаются более широким набором по внешнему виду, структуре, материалу, геометрическим размерам и т. д. Все это требует более сложных методов решения.

Характер внешних поверхностей. Наиболее простой случай – трехмерные объекты с плоскими гранями в виде различных многоугольников. Такие объекты математически хорошо описываются, и имеют более строгое решение. Другой тип – объекты с поверхностями в виде квадрик, к которым могут быть отнесены многие промышленно изготовленные объекты. Они также неплохо описываются одинаковыми примитивами с разными наборами параметров. Наконец, можно отметить еще один широкий класс объектов с произвольными поверхностями. Объекты этого типа представляют наибольшие трудности для распознавания. В этом случае может потребоваться дополнительное разбиение на подклассы, возможно, по иным признакам, для поиска подходящих методов решения.

Количество объектов на изображении. Задача распознавания усложняется с увеличением числа распознаваемых объектов. Наиболее часто решается задача распознавания для изображений с одним объектом. Кроме того, для изображений с несколькими объектами нужно различать случаи однородных (одинаковых) и разных объектов, которые могут различаться и методами решений. При наличии нескольких объектов, обычно, предварительно решается задача сегментации изображения.

Цель распознавания. Метод распознавания в значительной степени зависит и от цели распознавания. Примерами таких целей могут быть контроль изделий на конвейере, обнаружение препятствий для робота и определение направления передвижения, сортировки объектов и т. д. В каждом случае, в зависимости от цели распознавания, параметры задачи могут различаться в значительной степени и требовать соответствующих подходов к решению задачи.

Двумерные или трехмерные объекты. Эти два вида объектов имеют некоторые различия в методах обработки изображений и распознавании. Разнообразие в подходах могут вносить и форматы изображений (полутоновые, цветные, дальнометрические). Двумерные характерные признаки, найденные на изображении объекта, связаны с трехмерной моделью посредством видового преобразования. Обычно при распознавании проводят и процедуру правильности идентификации объекта.

10.6.2. Распознавание с использованием процедуры совмещения

В главе 4 рассматривались методы совмещения двумерных данных, которые подходят и для трехмерных данных. В общем случае алгоритм проверки соответствия множества точек изображения и множества точек трехмерной модели можно сформулировать следующим образом.

1. Высказывается гипотеза о соответствии между множествами точек.
2. На основании соответствия вычисляется преобразование, переводящее точки модели в точки изображения.
3. С помощью найденного преобразования вычисляются преобразованные модельные точки.
4. Преобразованные модельные точки сопоставляются с точками изображения, и делается вывод, подтвердилась ли гипотеза соответствия множества точек модели и изображения или нет.

Далее более подробно рассматриваются распознавания для двух случаев 3D–3D, 2D–3D.

Совмещение 3D–3D. Пусть трехмерные модели представлены в виде наборов характерных признаков, связанных с модельными точками. В качестве исходных данных обычно используются дальнометрические данные. Необходимо установить соответствие между парами точек «точка данных – модельный точечный признак». Соответствующей процедурой можно установить такое соответствие. Предполагается, что это соответствие включает в себя преобразования переноса и поворота.

Соответствие будет корректным, если правильно выбраны точки для проверки соответствия и не было ошибок в вычислениях. Для повышения надежности результата совмещения увеличивают число точек, участвующих в сопоставлении модели и изображения. На практике обычно используют до 10 точек.

После установления преобразования оно применяется ко всем точкам модели. Полученные результаты сравниваются с соответствующими точками изображения для проверки гипотезы соответствия. Если гипотеза не подтвердилась, то переходят к следующему возможному набору точек.

В соответствии с описанным подходом совмещения модели и изображения необходимо иметь эффективный алгоритм подбора пар точек. Для этого алгоритм должен учитывать характерные признаки и особенности конкретных объектов. Такими особенностями обычно бывают наличие углов, выступов, углублений, контрастных границ и т. п.

Если изображение не обладает достаточными особенностями, позволяющими сравнительно просто подбирать соответствующие пары точек «модель – объект», то необходимы более совершенные методы для повышения надежности распознавания.

Один из таких методов [12] заключается в том, что трехмерный объект представляется в виде каркасной модели с набором поворотных изображений. Поворотные изображения строятся по каркасной модели объекта и представляют собой двумерную гистограмму, построенную для выбранной вершины каркасной модели. При этом используется множество составляющих модельных точек. Объем, в котором находятся составляющие точки, зависит от двух параметров поворотного изображения. Первый параметр D представляет собой максимальное расстояние между выбранной вершиной каркаса и составляющей точкой. Вторым параметром A – это допустимый угол между нормалью в составляющей точке и нормалью в выбранной вершине. Кроме того, для каждой выбранной вершины каркасной модели вычисляются два значения расстояний – α и β . Расстояние α измеряется вдоль перпендикуляра к нормали поверхности, а β – вдоль перпендикуляра к касательной плоскости к данной вершине.

В начале работы алгоритма формируется накопительный массив $S[\alpha, \beta]$ с нулевыми значениями. После этого для каждой составляющей точки вычисляются параметры расстояний α и β по отношению к выбранной вершине, которые суммируются с соответствующими элементами массива $S[\alpha, \beta]$.

Поворотные изображения строятся для каждой вершины каркасной модели. Главное достоинство поворотных изображений состоит в отображении локальных свойств формы объекта в окрестности каждой вершины. После составления поворотных изображений выполняется сравнение двух множеств таких изображений. Для каждой пары поворотных изображений модели и объекта вычисляется коэффициент корреляции. Пары точек с наибольшими значениями коэффициентов корреляции берутся для сопоставления объектов.

Совмещение 2D–3D. Этот вариант совмещения заключается в сравнении трехмерного объекта и двумерного изображения. Задача осложняется не только наличием переноса и поворота объекта, но и за счет наличия перспективы на изображении. Полное преобразование определяется на основе множества соответствующих точек, или множества соответствующих отрезков, или эллипса на плоскости, окружности в пространстве и одной пары соответствующих точек. Иногда возможны комбинации характерных признаков всех трех типов.

Вычисленное преобразование между моделью и изображением позволяет применять его для характерных признаков трехмерной модели. В результате преобразования можно получить двумерные характерные признаки, которые могут быть обнаружены на изображении. При таком подходе может иметь место проблема экранирования на изображении характерных признаков другими объектами или поверхностями. Для решения этой проблемы можно применять известные алгоритмы удаления невидимых линий и поверхностей. Однако при этом существенно возрастает объем вычислений. Поэтому возникает необходимость использования более грубых алгоритмов или более мощных вычислительных средств, например, графические процессорные устройства, которые могут ускорить процесс вычисления на несколько порядков.

Совмещение объектов с гладкими поверхностями. Задача ставится так: имеется произвольный трехмерный объект с гладкими поверхностями и двумерное полутоновое изображение. Необходимо выполнить их сопоставление. Предлагается алгоритм, использующий модель на основе классов видимости. Трехмерный объект моделируется набором изображений объекта с разными ракурсами. Набор ракурсов содержит пять изображений: центральное изображение и 4 изображения, снятые с точек камеры, смещенной от центра влево, вправо, вверх и вниз.

Верхнее и нижнее изображения используются для вычисления кривизны краев объекта в направлении оси y . Левое и правое изображения используются для вычисления кривизны краев в направлении оси x . Аналогичные вычисления производятся для определения пространственного положения сгибов и контрольных точек карты перепадов. Получается так называемый 2,5-мерный ракурс модели, состоящий из трехмерных координат краев силуэта объекта, сгибов и контрольных точек, соответствующих центральному изображению с перепадами яркости. В представлении ракурса также хранится информация об x -, y -кривизне для каждой из этих точек. Этой информации достаточно, чтобы вычислить карту перепадов для любой другой точки из данного класса видимости.

В процессе обработки исходного изображения строится карта перепадов и формируется множество криволинейных сегментов, которые используются для индексирования базы данных при формировании гипотез «объект-вид». Сгенерированные гипотезы затем проверяются в процедуре сопоставления. Каждая гипотеза содержит идентификационные данные распознаваемого объекта и примерное его положение. Проверка гипотезы выполняется сглаживанием 2,5-мерного ракурса для каждой модели предполагаемого объекта относительно наблюдаемой карты перепадов.

При сопоставлении считается, что ракурс занимает центральное положение, и спроектированная карта перепадов этого ракурса модели сравнивается с наблюдаемой картой перепадов.

10.6.3. Распознавание на основе реляционных моделей

Трехмерные реляционные модели. Трехмерные реляционные модели состоят из трехмерных примитивов и пространственных отношений. В качестве трехмерных примитивов могут использоваться простые трехмерные объемы и поверхности, характерные признаки в виде отрезков и сегментов. Широкое применение в реляционных моделях имеют обобщенные цилиндры. Наряду с ними в реляционные модели включаются разновидности трехмерного отношения связности.

Другой подход в построении реляционных моделей основан на использовании набора стандартных объемных примитивов из трех типов: стержни, пластины и капли. В этом случае реляционная модель состоит из этих объемных примитивов, связанных между собой пространственными отношениями. Главная особенность указанных объемных примитивов заключается в том, что стержни имеют основной одномерный параметр, пластина – двумерный параметр, а капля – трехмерный.

Описание этих примитивов включает в себя следующие элементы.

Стержень описывается двумя логическими концевыми точками, логическим множеством внутренних точек и логическим центром тяжести.

Пластина имеет множество краевых точек, множество точек поверхности и центр тяжести. У **капли** множество точек поверхности и центр тяжести.

Пространственные отношения реляционной модели описывают структуру объекта. При этом могут быть указаны отношения параллельности, перпендикулярности примитивов, пространственного положения по одну сторону или разные относительно каких-то поверхностей или других примитивов, а также связность примитивов.

Реляционные модели на основе классов видимости. Данный метод используется в тех случаях, когда исходные данные представлены двумерными изображениями. Вместо истинных трехмерных моделей используются модели на основе классов видимости. Перед построением моделей на основе классов видимости проводится обучение системы. Исходной базой для обучения могут быть искусственно сформированные данные или набор реальных изображений объектов. Множество характерных двумерных признаков, выделяемых на изображении, зависит от

класса объекта. Признаки, выделяемые на каждом изображении из учебного набора, используются для подготовки реляционного описания соответствующего вида объекта. Затем выполняется формирование классов видимости объекта на основе кластеризации реляционных описаний.

Каждый класс видимости представляется комбинированным реляционным описанием, содержащим характерные признаки, обнаруженные на всех видах этого класса видимости. Реляционной моделью класса видимости является комбинированное реляционное описание. Для каждого объекта, как правило, строится порядка пяти классов видимости с собственными реляционными описаниями. В зависимости от числа моделей в базе данных возможен полный перебор реляционных сопоставлений или же выполняется предварительная реляционная индексация.

Распознавание на основе функциональных моделей. Функциональные модели представляют собой классы, каждый из которых содержит модели, имеющие одинаковое функциональное назначение. В качестве примера принципов распознавания с использованием функциональных моделей здесь будет описана универсальная система распознавания объектов с использованием характеристик формы и функциональных свойств [13].

В этой системе выделены три уровня знаний:

- примитивы знаний для построения функциональных описаний;
- определение каждой категории с точки зрения функционального назначения;
- иерархия категорий всех объектов в базе знаний.

Примитивы знаний являются параметризованной процедурой, которая реализует базовое понятие, связанное с геометрическими, физическими или функциональными свойствами. На вход примитива знаний поступает часть описания пространственной формы, а на выходе – значение, обозначающее, насколько хорошо входные данные соответствуют заданным требованиям.

В описываемой универсальной системе распознавания определены 6 примитивов знаний, предназначенных для реализации следующих понятий:

- относительная ориентация – определяет, насколько хорошо относительное расположение двух поверхностей соответствует некоторому желательному расположению;
- размеры – проводит проверку шести геометрических размеров (ширины, высоты, глубины, площади, непрерывной поверхности, объема);

-
- близость – определяет качественное пространственное отношение между элементами формы объекта;
 - устойчивость – проверяет устойчивость объекта в конкретных условиях;
 - отсутствие препятствий – проверяет отсутствие препятствий в указанном объеме пространстве между частями объекта;
 - огораживание – проверяет наличие требуемых полостей и вогнутостей объекта.

Функциональные свойства определяют классы объектов по функциональному назначению. Например, классы объектов мебель, посуда, инструменты, электрические приборы и т. д. Для этих классов определяют некоторые шаблоны, позволяющие выполнять проверку на функциональность каждого объекта класса. Для класса мебель, например, может быть определен шаблон на проверку устойчивости шкафа, для стула – шаблон на наличие поверхности для сидения и т. д.

Иерархия категорий определяет иерархическое дерево ветвления для категорий классов. Например, класс посуда может делиться на чашки, тарелки, кастрюли, графины и т. д. В свою очередь, каждая такая категория (чашки, тарелки, кастрюли, графины) может состоять из конкретных объектов типа чашка для чая, для кофе, тарелка глубокая, десертная мелкая и т. д.

Процесс распознавания делится на два этапа: предварительный этап и этап распознавания. На первом этапе все объекты обрабатываются одинаково, независимо от их категорий. Он включает в себя анализ трехмерных дальнометрических данных, и среди них перебираются все возможные функциональные элементы. Данные, полученные на предварительном этапе, затем используются на этапе распознавания для индексации и упорядочивания категорий объектов. В индекс включаются функциональный элемент, а также его площадь и объем. Те категории, которые невозможно использовать для сопоставления на основе данных индекса, из рассмотрения исключаются.

Остальные категории упорядочиваются и обрабатываются для получения оценки соответствия. Для получения такой оценки для каждой гипотезы о классе объектов вызываются примитивы знаний, которые вычисляют степень соответствия функционального элемента, найденного в исходных данных, требованиям данного класса. Степень соответствия имеет количественную оценку, составленную из количественных значений, полученных от примитивов знаний.

Распознавание по внешнему виду объектов. Распознавание по внешнему виду объекта отличается достаточно простой процедурой рас-

познавания. Она заключается в том, что для каждого распознаваемого объекта составляется набор двумерных изображений, каждое из которых затем сравнивается с входным изображением. Как правило, сравнение выполняется в виде полутоновых изображений. Простота процедуры распознавания позволяет вести обработку с большими наборами изображений, что, естественно, повышает качество распознавания. Наибольших успехов распознавания данным методом достигнуто в задаче распознавания человеческих лиц.

Распознавание состоит из двух этапов обработки. Первый этап – обучение системы, на котором создается база данных маркированных изображений, состоящих из эталонных изображений и их меток. На этапе распознавания неизвестный объект распознается сравнением его изображения с изображениями из базы данных. Объекту присваивается метка того эталонного изображения, которое наиболее соответствует распознаваемому объекту. Наиболее похожее эталонное изображение можно найти из условия минимума евклидова расстояния между изображениями или из условия максимума скалярного произведения между ними.

Для успешного решения задачи распознавания, естественно, предъявляется ряд требований к эталонным изображениям. Например, для распознавания человеческого лица нужно иметь набор изображений с разными выражениями лица, поворотами головы, вариантами освещения.

Предъявляются требования и для объекта распознавания – объект должен быть отчетливо ограничен, иметь примерно одинаковые размеры и положения. При решении задачи необходимо учитывать и наличие фона, который должен учитываться системой на этапе обучения.

Важной задачей повышения качества распознавания является проблема уменьшения количества сигнальных признаков, используемых для представления объектов. С этой целью используются различные базисы пространства изображений, а также представление изображений в виде суммы особых изображений, таких как элементарные края, волнообразная рябь и т. д. При этом представление изображений линейной комбинацией изображений ортонормированного базиса позволяет вычислить энергию этого изображения как сумму квадратов коэффициентов базиса.

Базисные изображения для набора эталонных изображений. Будем считать, что можно найти ортонормированное базисное множество изображений B размером $N=R \times C$, обладающее следующими свойствами.

1. $B = \{F_1, F_2, \dots, F_m\}$, где $m \ll N$.

2. Для всех M изображений I_j из эталонного множества выполняется

$$I_j^m = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m$$

и

$$\sum_{j=1}^m \left(\frac{\|I_j^m - I_j\|^2}{\|I_j\|^2} \right)^2 < P,$$

где I_j^m – аппроксимация исходного изображения I_j в виде линейной комбинации m базисных изображений, а P – некоторый процент, являющийся порогом удовлетворительного среднего качества представления набора изображений. Практически этот порог может иметь значение около 5%.

Соотношение $m \ll N$ при работе с базисным набором B позволяет снизить требования по затратам памяти и ускорить вычисления. Базисные изображения F_i называются главными компонентами множества эталонных образцов. Алгоритм распознавания может быть построен в виде двух этапов следующим образом.

На этапе обучения вычисляется среднее по эталонным изображениям. Затем выполняется генерация множества разностных изображений D , каждое из которых является разностью между эталонным и усредненным изображением. Разностные изображения D можно представить вектором из N элементов в виде двумерного массива $N = R \times C$.

Далее вычисляется ковариационная матрица эталонных изображений Σ_D . По определению, элемент этой матрицы $\Sigma_D[i, i]$ равен дисперсии i -го пикселя, элемент $\Sigma_D[i, j]$ – ковариации i -го и j -го пикселей по всем эталонным изображениям. Ковариационная матрица вычисляется так:

$$\Sigma_D = D^T D.$$

Размеры этой матрицы очень велики, $N \times N$, т.е. в минимальном варианте 256×256 или 512×512 . Поэтому вместо ковариационной матрицы вычисляют связанную матрицу

$$\Sigma^*_D = D D^T,$$

размеры которой существенно меньше и равны $m \times m$. Связь собственных векторов и собственных значений этих двух матриц выражается следующими соотношениями:

$$\begin{aligned} \Sigma_D F &= \lambda F; \\ \Sigma^*_D F^* &= \lambda F^*; \\ F &= D^T F^*, \end{aligned}$$

где λ – вектор собственных значений матрицы Σ_D , F – вектор собственных векторов матрицы Σ_D ;

F^* – вектор собственных векторов матрицы Σ^*_D .

Заключение

Учебное пособие представляет собой руководство по компьютерному зрению. В нем содержится теоретический материал и примеры, которые могут быть полезны для широкого круга специалистов в области компьютерных технологий. Особый интерес представляют развивающиеся области компьютерного зрения по работе с базами данных изображений, созданию систем виртуальной и дополненной реальности, автоматизации обработки изображений, в том числе в получении и обработке медицинских изображений, создании компьютерной картографии и геоинформационных систем.

Материал, изложенный в пособии, может быть положен в основу разработки пакетов программ для решения широкого круга задач компьютерного зрения, а также для проведения исследований по разработке новых эффективных методов их решения на основе современных информационных технологий.



Библиографический список

1. *Шапиро, Л.* Компьютерное зрение / Л. Шапиро, Дж. Стокман ; пер. с англ. – М. : БИНОМ. Лаборатория знаний, 2006. – 752 с.
2. *Визильтер, Ю. В.* Обработка и анализ изображений в задачах машинного зрения : курс лекций и практических занятий / Ю. В. Визильтер [и др.]. – М. : Физматкнига, 2010. – 672 с.
3. *Лукьяница, А. А.* Цифровая обработка изображений / А. А. Лукьяница, А. Г. Шишкин. – М. : Ай-Эс-Эс Пресс, 2009. – 518 с.
4. *Форсайт, Д. А.* Компьютерное зрение. Современный подход / Д. А. Форсайт, Ж. Понс ; пер. с англ.. – М. : Вильямс, 2004.
5. *Haralick, R. M.* A measure of circularity of digital figures // IEEE Trans. Syst., Man, and Cybern. – 1974. – V. SMC-9. – P. 62–66.
6. *Shi, J.* Normalized cuts and image segmentation / J. Shi, J. Malik // IEEE Conf. Comput. Vision and Pattern Recog. – 1997. – P. 731–737.
7. *Селянкин, В. В.* Анализ и обработка изображений в задачах компьютерного зрения : учеб. пособие / В. В. Селянкин, С. В. Скороход. – Таганрог : Изд-во ЮФУ, 2015. – 82 с.
8. *Селянкин, В. В.* Решение задач компьютерного зрения : учеб. пособие. – Таганрог : Изд-во ЮФУ, 2016. – 92 с.
9. *Haralick, R.* The consistent labeling problem / R. Haralick, L. Shapiro // I. IEEE Trans. – 1979. – V. PAMI-1. – P. 173–184.
10. *Haralick, R.* The consistent labeling problem / R. Haralick, L. Shapiro // I. IEEE Trans. – 1979. – V. PAMI-2. – P. 193–203.
11. *Селянкин, В. В.* Компьютерная графика : учеб. пособие. – Таганрог : Изд-во ТТИ ЮФУ, 2009. – 212 с.
12. *Johnson, A. E.* Efficient multiple model recognition in cluttered 3-D scenes. Proc. / A. E. Johnson, M. Hebert // IEEE Conf. Comput. Vision and Pattern Recognition. – 1998. – P. 671–677.
13. *Stark, L.* Generic Object Recognition Using Form and Function / L. Stark, K. Bowyer. – Singapore : World Scientific Publishing Co. Pte. Ltd., 1996.

Владимир Васильевич СЕЛЯНКИН
КОМПЬЮТЕРНОЕ ЗРЕНИЕ
АНАЛИЗ И ОБРАБОТКА ИЗОБРАЖЕНИЙ

Учебное пособие
Издание второе, стереотипное



Зав. редакцией литературы по информационным технологиям и системам связи *О. Е. Гайнутдинова*



ЛР № 065466 от 21.10.97
Гигиенический сертификат 78.01.10.953.П.1028
от 14.04.2016 г., выдан ЦГСЭН в СПб
Издательство «ЛАНЬ»
lan@lanbook.ru; www.lanbook.com;
196105, Санкт-Петербург, пр. Юрия Гагарина, 1, лит. А.
Тел.: (812) 412-92-72, 336-25-09.
Бесплатный звонок по России: 8-800-700-40-71

Подписано в печать 14.05.21.
Бумага офсетная. Гарнитура Школьная. Формат 84×108^{1/32}.
Печать офсетная. Усл. п. л. 7,98. Тираж 50 экз.
Заказ № 636-21.

Отпечатано в полном соответствии
с качеством предоставленного оригинал-макета
в АО «Т8 Издательские Технологии».
109316, г. Москва, Волгоградский пр., д. 42, к. 5.