

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

А. Ю. Тропченко, А.А. Тропченко
МЕТОДЫ ВТОРИЧНОЙ ОБРАБОТКИ И
РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ
Учебное пособие

 УНИВЕРСИТЕТ ИТМО

Санкт-Петербург

2015

Тропченко А.А., Тропченко А.Ю. Методы вторичной обработки и распознавания изображений. Учебное пособие. – СПб: Университет ИТМО, 2015. – 215 с.

Учебное пособие охватывает основные методы теории цифровой обработки сигналов, используемые при вторичной обработке изображений, а также основы теории распознавания образов. В пособие включены и сведения о используемых в настоящее время методах сжатия и восстановления изображений, которые составляют отдельную важную группу методов обработки изображений. Материал пособия разбит на три раздела, включая введение. В каждом разделе, кроме введения, приведены краткие теоретические сведения об используемых для решения соответствующих задач математических методах.

Пособие может быть использовано при подготовке магистров по направлениям 09.04.04 «ПРОГРАММНАЯ ИНЖЕНЕРИЯ», 09.04.02 «ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА», а также инженеров и аспирантов.

Рекомендовано к печати ученым советом факультета Компьютерных технологий и управления, протокол №6 от 23 июня 2015 г.

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2015

©А.Ю. Тропченко,

А.А. Тропченко, 2015

СОДЕРЖАНИЕ

Введение	6
1. МЕТОДЫ ВТОРИЧНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ...	10
1.1. Задача сегментации изображений	11
1.2. Методы сегментации изображения (локализации объекта)	15
1.2.1. Метод порогового ограничения	17
1.2.2. Операции над бинарными изображениями	35
1.2.3. Метод наращивания областей	37
1.3. Методы выделения границ объектов	43
1.3.1. Градиентный метод	44
1.3.2. Выделение контуров	45
1.3.3. Градиентная обработка при линейной фильтрации	46
1.4. Описание контура объекта	47
1.5. Скелетизация объекта	48
1.6. Выделение характерных точек объекта	49
2. ОСНОВЫ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ	52
2.1. Задача распознавания изображений	54
2.2. Методы распознавания на основе сравнения с эталоном	60
2.2.1. Распознавание по проекциям	68
2.2.2. Распознавание на основе двумерных гистограмм	69
2.3. Признаковые методы распознавания	71
2.3.1. Построение решающих функций для двоичных наборов параметров	79
2.4. Контурные методы распознавания	81
2.4.1. Определение расстояния между описаниями объектов	84
2.5. Вероятностный подход к распознаванию	86
2.6. Структурные методы распознавания	88
2.6.1. Непроизводные структурные элементы	88
2.6.2. Составные структурные элементы	89
2.7. Методы распознавания, основанные на знаниях	94
2.8. Критерии эффективности распознавания изображений	96
2.9. Задача распознавания лиц	104
2.9.1. Метод главных компонент	104
2.9.2. Линейный дискриминантный анализ	108

2.9.3. Синтез объектов линейных классов	109
2.9.4. Гибкие контурные модели лица	110
2.9.5. Сравнение эластичных графов	111
2.9.6. Методы, основанные на геометрических характеристиках лица	114
2.9.7. Сравнение эталонов	115
2.9.8. Оптический поток	116
2.9.9. Скрытые Марковские модели	118
3. ПРИКЛАДНЫЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ.....	125
3.1. Классификация и применение искусственных нейронных сетей.....	130
3.1.1. Основные классы решаемых задач при распознавании человека по изображению лица	130
3.1.2. Поиск изображения в больших базах данных	131
3.1.3. Задача контроля доступа	132
3.1.4. Задача контроля фотографии в документах	133
3.1.5. Нейросетевые методы распознавания человека по изображению лица	134
3.1.6. Модель искусственного нейрона.....	137
3.1.7. Классификация нейронных сетей.....	141
3.2. Разделение пространства признаков на области и извлечение ключевых признаков	143
3.2.1. Многослойные нейронные сети	143
3.2.2. Нейронные сети высокого порядка и моментные НС	152
3.2.3. Радиально-базисные нейронные сети	153
3.3. Топологически упорядоченное преобразование пространства	155
3.4. Распознавание с учётом топологии пространства	158
3.4.1. Когнитрон	159
3.4.2. Неокогнитрон.....	159
3.4.3. Свёрточные нейронные сети	161
3.5. Достоинства и недостатки нейросетевыхметодов распознавания.....	163
3.6. Применение ИНС для извлечения ключевых характеристик.....	164
3.6.1. Применение ИНС для извлечения ключевых характеристик лица.....	164
3.7. Применение ИНС для классификации образов	165
3.7.1. Применение ИНС для классификации напрямую по входным сигналам.....	166
3.7.2. Системы распознавания по входным сигналам с	

неявным извлечением ключевых характеристик внутри сети.....	175
3.7.3. Применение ИНС для классификации по заранее извлеченным признакам.....	178
3.7.3.1. Метод собственных фильтров для распознавания лица.....	178
3.7.3.2. Метод сравнения эластичных графов на основе вейвлетов Габора.....	179
3.7.3.3 Методы распознавания на основе Скрытых марковских моделей.....	181
3.7.3.4.Эксперименты по распознаванию лица.....	183
3.8. Аппаратные и программные средства реализации ИНС...	186
3.8.1 Программное обеспечение для моделирования ИНС.....	186
3.8.2. Аппаратная реализация нейронных сетей.....	190
3.8.3. Принципы построения нейросистем на базе нейрочипа.....	199
Заключение.....	203
Литература.....	204

Введение

В развитии систем технического зрения наметилось два направления. Первое из них связано с созданием универсальных исследовательских комплексов обработки изображений в различных технологических ситуациях. Такие комплексы снабжаются различными видео датчиками, аппаратурой ввода-вывода видеоинформации в ЭВМ большой мощности, имеют развитую систему программного обеспечения, разнообразные периферийные средства. Назначение таких систем — исследование и испытание алгоритмов обработки изображений для различных технологических ситуаций, а также автоматизированное проектирование архитектур и схемотехники вычислителей, реализующих полученные алгоритмы [1,7,13].

Второе направление — создание специализированных систем технического зрения для конкретных роботов [46]. Исходной информацией при этом могут служить результаты предварительного моделирования системы на универсальном комплексе либо априорные сведения о характере входных изображений и эвристические соображения об алгоритмах их обработки.

В последние годы значительно возрос интерес к электронным, цифровым и оптическим методам обработки изображений с целью повышения их качества. Широкое освещение получили работы, связанные с космическими и биомедицинскими исследованиями. Из числа других применений следует упомянуть аэрофотосъемку и промышленную радиографию.

Повышение качества изображений достигается двумя видами обработки изображений: реставрацией (исправлением) изображений и их улучшением [13]. Под реставрацией обычно понимают процедуру восстановления или оценивания элементов изображения, целью которой является коррекция искажений и наилучшая аппроксимация идеального неискаженного изображения. Для улучшения изображений используется комплекс операций, призванных улучшить восприятие изображения наблюдателем или же преобразовать его в другое изображение, более удобное для машинной обработки.

Процедура улучшения изображений сводится к выполнению комплекса операций с целью либо улучшения визуального восприятия изображения, либо преобразования его в форму, более удобную для визуального или машинного анализа. В системах улучшения изображений не делается попытки приблизить воспроизводимое изображение к некоторому идеализированному оригиналу, такая задача решается при реставрации изображений. Известны случаи, когда искаженное изображение субъективно

воспринимается лучше, чем неискаженный оригинал. Примером может служить изображение с подчеркнутыми границами (контурами).

При машинной обработке улучшение изображения тесно связано с задачей извлечения информации. Пусть, например, система улучшения изображений производит подчеркивание границ исследуемого изображения путем высокочастотной фильтрации. Обработанное изображение затем вводится в ЭВМ, которая выполняет операцию прослеживания контура объекта, определяет его форму и размеры. В этом примере система улучшения изображений используется для того, чтобы подчеркнуть важнейшие признаки исходного изображения и, следовательно, облегчить задачу извлечения информации.

Сфера применения систем технического зрения непрерывно расширяется, что требует, с одной стороны, разработки все новых методов и алгоритмов обработки изображений, а с другой — теоретического обобщения полученных результатов.

Проблема обработки изображений включает в себя такие, ставшие уже почти классическими, задачи, как сегментация, фильтрация помех и выделение изображений из фона, определение границ объектов, распознавание образов. Центральное место среди названных понятий занимает задача распознавания образов [15,23,45]. Решение ее техническими средствами может быть осуществлено путем моделирования операций, выполняемых живыми организмами в процессе коммуникации и восприятия окружающего мира. Эти способности достаточно хорошо изучены на различных животных (киты, дельфины, летучие мыши, некоторые приматы). Однако наиболее естественно положить в основу модели распознавания способности человека и его реакции на окружающую действительность.

Распознавание образов как научное направление включает в себя большое число различных дисциплин и использует методы, характерные для каждой из них. Наряду с вездесущей информатикой среди них можно выделить прикладную физику, необходимую при разработке эффективных датчиков, а также различные разделы математики, лежащие в основе методов обработки данных. Распознавание образов есть совокупность методов и средств, позволяющих, по меньшей мере, достигнуть, а если возможно, то и превзойти естественные средства восприятия и анализа окружающего мира живыми существами [45].

Получение символического описания изображений представляет собой задачу перехода от набора простейших признаков изображения, таких, как значения яркости, контурные точки или параметры текстуры, к значительно меньшему набору средств

описания, которые могут служить в качестве исходных данных для последующей семантической интерпретации. Типичными графическими символами являются цепочки контурных точек, образующих границу объекта, связанные области постоянной яркости, цвета или текстуры и элементарные фигуры, такие, как прямоугольники, окружности, треугольники.

К сожалению, не существует общепринятого набора визуальных символов, которые необходимы и достаточны для описания изображения. Отсутствие набора единых визуальных символов создает определенные трудности при анализе изображений. Во-первых, существует проблема отбора, суть которой состоит в том, чтобы определить, какие символы необходимо сформировать из признаков изображения для решения конкретных задач анализа изображений. Кроме того, имеется проблема определения необходимой точности при формировании символов.

Основной этап при формировании символического описания изображения по массиву элементов или набору простейших признаков заключается в определении геометрических соотношений и связанности между элементами, относительно которых предполагается, что они принадлежат одному классу.

Большая часть документации промышленных изделий, снимков, различных карт в том или ином виде содержат тексты, буквенные и цифровые обозначения, специальные символы, которые подлежат вводу наряду с графической и визуальной информацией и, что более важно, составляют неотъемлемую часть всего изображения. Поэтому полная и правильная интерпретация изображения возможна только при правильном “чтении” буквенных текстов, цифр, символов. Тексты и символы могут быть воспроизведены типографским способом, на пишущей машинке, написаны по трафарету или от руки. При их “чтении” можно выделить две самостоятельные задачи:

- отделение символьной информации от графической и визуальной информации;
- распознавание и интерпретация каждого символа.

Вторая задача породила целую серию подзадач, начиная от считывания и распознавания машинописных цифр до распознавания рукописных знаков (буквы и цифры), символов (например, обозначение месторождений полезных ископаемых на картах), нотных записей и слитно написанных от руки текстов.

О сложности распознавания рукописных знаков свидетельствует тот факт, что даже человек, читая их при отсутствии контекста, делает около 4 % ошибок [15]. Причины ошибок кроются в бесконечном разнообразии вариаций формы знаков, которые обусловлены наличием ограничений в начертании знаков и символов или их отсутствием, стилем, образованием, опытом, настроением,

здоровьем и другими характеристиками пищевого, а также в качестве пищевого инструмента и поверхности, на которой едят, методах считывания и алгоритмах распознавания.

1. МЕТОДЫ ВТОРИЧНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

Задачу системы технического зрения можно определить как процесс выделения, идентификации и преобразования информации, полученной из изображений. Этот процесс можно разделить на шесть основных этапов: снятие информации, предварительная обработка информации, сегментация, описание, распознавание, интерпретация [15].

Снятие информации представляет собой процесс получения визуального изображения. Предварительная обработка информации заключается в использовании таких методов как понижение шума или улучшение изображения отдельных деталей. Сегментация — процесс выделения на изображении интересующих объектов. При описании определяются характерные параметры (например, размеры или форма), необходимые для выделения требуемого объекта на фоне других. Распознавание представляет собой процесс идентификации объектов. Интерпретация выявляет принадлежность к группе распознаваемых объектов.

Таким образом, обработка изображений является многоэтапной процедурой, причем каждому этапу соответствует изображение того или иного вида (см. рис.1.1).

Все подлежащие обработке изображения могут быть подразделены на 4 типа [1,7]:

- 1 - многоуровневые или полутоновые изображения;
- 2 - бинарные изображения или битовые карты изображений;
- 3 - изображения характерных линий контуров или скелетов объектов на изображении;
- 4 - изображения - наборы характерных точек - координат углов, точек пересечения линий, изменения кривизны линии и т.д.

Чаще всего исходные изображения представлены в форме изображений типа 1. Очевидно, что при последовательном переходе от изображения 1 к изображению 4 упрощается его описание, что облегчает в дальнейшем его распознавание. По сути, при работе с изображениями типов 3 и 4 работа выполняется с некоторыми наборами признаков, описывающих исходное изображение объекта (см. табл. 1.1).

При этом первые три типа преобразований характерны для задач обработки и распознавания изображений, а три последних - для задач синтеза изображений.

Задача подавления помех входит, с одной стороны, в задачу улучшения изображения, а с другой — ее можно рассматривать как часть задачи сегментации. Конечной целью сегментации изображений является разбиение поля зрения на области объектов и область фона, что является затруднительным, т.к. понятия “объект”, “фон”, “граница объекта” весьма условны. Методы и алгоритмы сегментации можно

рассматривать как формализацию понятия выделяемости объекта из фона.

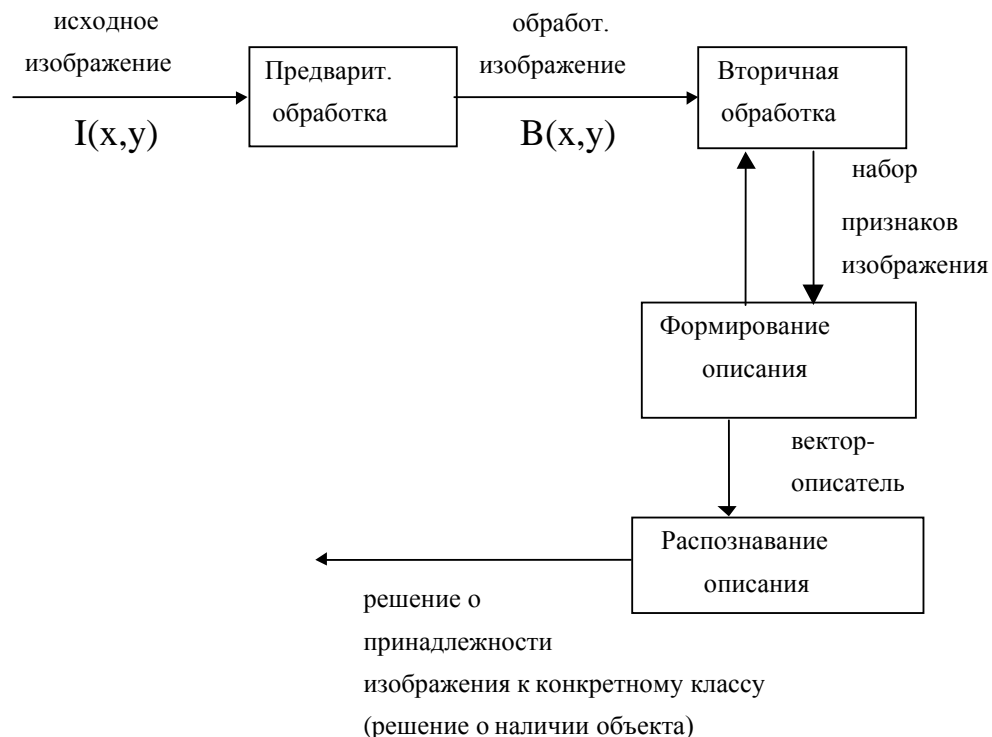


Рис.1.1. Основные этапы обработки изображений

Таблица 1.1.

Типы преобразований

Преобразование	Используемые процедуры
1 → 2	Сегментация и бинаризация
2 → 3	Векторизация, построение контура и скелета
3 → 4	Векторизация, сегментация кривых, определение экстремумов
4 → 3	Интерполяция и аппроксимация
3 → 2	Заполнение контура
2 → 1	Низкочастотная фильтрация и рандомизация

1.1. Задача сегментации изображений

Объект изображения (образ) можно рассматривать как совокупность областей интенсивности. Для выделения области интенсивности на изображении можно использовать не только методы распознавания образов, но и методы разбиения изображений или методы сегментации областей. Алгоритмы сегментации можно рассматривать как формализацию понятия выделяемости объекта из фона.

Сегментацией называется процесс подразделения сцены на составляющие части или объекты. Сегментация является одним из

основных элементов работы автоматизированной системы технического зрения, т.к. именно на этой стадии обработки объекты выделяются из сцены для дальнейшего распознавания и анализа. Сегментация изображения представляет собой разделение или разбиение изображения на области по сходству свойств их точек [1,13,34]. Наиболее часто сегментацию проводят по яркости для одноцветного изображения и цветовым координатам для цветного изображения.

Конечной целью сегментации изображений является разбиение поля зрения D на области объектов D_1, \dots, D_s и область фона D_f .

Обработка изображения – это любая форма обработки информации, для которой входные данные представлены изображением. Обработка изображений может осуществляться как для получения изображения на выходе (например, предобработка для каких-либо других задач), так и для получения другой информации (например, выделение объектов, подсчет объектов, распознавание образов и так далее).

Цифровая обработка изображений, в отличие от аналоговой, которая еще совсем недавно была очень популярна, является более надежным, точным и гибким способом обработки изображения. Для цифровой обработки активно применяются вычислительные машины, используя графические системы. И, естественно, от развития и уровня вычислительных систем зависят возможности человека в решении задач, связанных с обработкой изображений.

На сегодняшний день существует множество средств и программных систем для обработки изображений. Большинство из них являются узконаправленными, но тем ни менее общий круг рассматриваемых ими вопросов очень велик. Это и основные методы фильтрации, и вейвлет-преобразования, и методы улучшения, восстановления и сжатия изображений и так далее. Все эти методы обработки и другие активно применяются в повседневной жизни. Например, в медицине, в военной области, в машинном зрении, в системах, основанных на биометрических данных человека (сканирование сетчатки, радужной оболочки, отпечатков пальцев и так далее), в задачах сегментации изображений.

Одним из методов обработки изображений, о котором пойдет речь ниже, является пороговая обработка. Этот метод занимает одно из центральных мест в прикладных задачах сегментации изображений. Под сегментацией изображений понимают разделение исходного изображения на составляющие его области или объекты [13]. Такой подход в обработке изображений на сегодняшний день нашел очень широкое применение.

Сама задача сегментации изображений представляет собой одну из самых сложных задач обработки изображений. Конечный

результат зачастую определяется точностью сегментации, поэтому при выборе того или иного метода сегментации нужно уделять большое внимание надежности алгоритма [1]. Однако, единого, общепризнанного подхода, который бы лежал в основе большинства алгоритмов, нет. Нет также и единого алгоритма, который позволял бы проводить приемлемую сегментацию для любого изображения. В этом и заключается одна из сложностей сегментации, и это является причиной большого числа различных подходов при решении данных задач обработки изображений.

В задачах обработки изображения входными данными является некоторое изображение. На выходе может быть не только обработанное изображение, но и некоторая его характеристика. Изображения бывают различными: от бинарных, то есть тех, в которых присутствует лишь два цвета (как правило, черный и белый), до 32-х битных изображений. Остановимся на самых распространенных 24-х битных изображениях формата .bmp (от англ. bitmap). Этот формат предназначен для хранения растровых изображений, то есть изображений, представляющих собой сетку пикселей или же точек цветов. Помимо него есть, несомненно, множество других форматов, таких как .jpeg (.jpg), .png, .tiff и другие. Отличаются они, как правило, способом кодировки изображения. Например, формат .jpeg содержит сжатые данные, которые были получены по одноименному алгоритму JPEG, являющемуся алгоритмом сжатия с потерей качества. Чтобы рассматривать изображение без потери качества и с полной глубиной цвета был выбран формат .bmp.

Изображение является 24-х битным, так как содержит в своем цвете три основных составляющих: R (красный), G (зеленый) и B (синий). Каждый из цветов варьируется в пределах от 0 до 256, то есть в пределах: $[0; 2^8]$. Таким образом, всего цветов может быть $2^8 \cdot 2^8 \cdot 2^8 = 2^{24}$, или же 16.777.216. Цветовое пространство RGB на сегодняшний день очень распространено. Подавляющее большинство дисплеев телефонов, плееров, GPS-навигаторов, цифровых фотоаппаратов и видеокамер отображают именно 16 миллионов цветов. На самом деле компьютеры обрабатывают не сами изображения, а только массивы дискретных чисел [7]. Изображение в таком представлении, вообще говоря, – это двумерный массив точек, то есть матрица, каждый элемент которой есть набор трех значений: R, G и B. Формально изображение можно записать как некоторую функцию.

$$f[x, y, p(x, y)] \quad (1.1)$$

где x, y – координаты пикселя изображения, а $p(x, y)$ – яркость

пикселя в точке с координатами (x, y) . Яркость пикселя определяется как:

$$p(x, y) = 0.3 * R(x, y) + 0.59 * G(x, y) + 0.11 * B(x, y) \quad (1.2)$$

где $R(x, y)$, $G(x, y)$ и $B(x, y)$ – значения составляющих R , G и B соответственно в точке (x, y) . В компьютерной реализации эта матрица представлена в виде сплошного массива, записанного по строкам изображения, каждый элемент которого является структурой RGB.

Часто для анализа яркости изображения строят его гистограмму. Гистограмма изображения – это гистограмма уровней яркости изображения. Она позволяет оценить количество и разнообразие оттенков изображения, а также его общий уровень яркости. Гистограмма – это функция, аргументом которой являются всевозможные уровни яркости одной из составляющих R, G или B (или же непосредственно яркость, вычисленная по формуле (1.1)), то есть область определения лежит в диапазоне $[0; 255]$, а значением является количество точек, которые принимают данный уровень яркости. Для работы с изображениями на компьютере с использованием языков программирования высокого уровня существует достаточно много средств. Самыми распространенными являются: Open GL (Open Graphics Library) – это открытая графическая библиотека, DirectX – набор функций, разработанных для решения задач, связанных с видеопрограммированием под Microsoft Windows, а также непосредственно Microsoft Visual Studio, которая также представляет множество встроенных средств для работы с графикой.

Понятие сегментации, данное выше, является обобщенным понятием. Вообще говоря, изображение для наблюдателя часто представлено в виде некоторых однородных участков, отличающихся друг от друга различными характеристиками. Количество таких типов (или же классов) обычно невелико. Все изображение можно разбить на некоторое количество непересекающихся областей, каждая из которых является изображением одного из типов (классов). При анализе таких изображений целью любой системы является определение этих областей и указания их номера типа. Обработка изображения, позволяющая получить такую совокупность сведений о нем, и называется сегментацией [13]. Иными словами, предполагается, что области изображения соответствуют реальным объектам или же их частям.

Однако существуют изображения, в которых вся картина разбита на области, не отличающиеся друг от друга ни по каким

характеристикам. Тогда вся информация представляет в данном случае совокупность границ между этими областями. Простой пример: кирпичная или плиточная кладка.

Методы сегментации изображений делятся на два класса:

- автоматические, то есть такие методы, которые не требуют взаимодействия с пользователем;

- интерактивные или же ручные методы, использующие введенные пользовательские данные во время работы.

Задача сегментации изображения, как правило, применяется на некотором этапе обработки изображения, чтобы получить более точные и более удобные представления этого изображения для дальнейшей работы с ним.

1.2. Методы сегментации изображения (локализации объекта)

Методы и алгоритмы сегментации можно рассматривать как формализацию понятия выделяемости объекта из фона.

Надежность алгоритмов сегментации зависит от того, насколько точно и полно при этом учтена дополнительная информация, которая в основном состоит из следующих сведений [13,16]:

- число объектов S ;
- некоторые характеристики распределения яркости в областях объектов или фона, например экстремальные значения яркости, количество перепадов яркости;
- оценки яркостного перепада при переходе из области фона в область объектов;
- форма объекта;
- информация о том, какую часть поля зрения занимает объединение областей объектов.

Методов сегментации существует великое множество, и разные методы ориентированы на разные свойства разбиения изображения. Поэтому при выборе метода сегментации в той или иной задаче следует руководствоваться тем, какие же свойства разбиения действительно важны и какими свойствами обладает исходное изображение. Также необходимо решить, какая степень детализации, до которой доводится разделение на классы, оказывается приемлемой. Все зависит от каждой конкретной решаемой задачи. Например, при анализе микросхем задачей выделения объектов может быть выделение блоков микросхем и радиодеталей, а может быть обнаружение трещин на этих радиодеталях. Тогда логично, что в первом случае необходимо ограничиться более крупной детализацией.

Перейдем теперь к конкретным методам сегментации. Они делятся на методы разметки точек и выделения границ. Метод разметки точек делится, в свою очередь, на метод порогового ограничения по яркости и метод наращивания областей, причем в системах технического зрения наиболее распространен метод порогового ограничения.

Алгоритмы сегментации также можно разделить исходя из их свойств на два типа [13, 16]:

- 1) основанные на базовом свойстве яркости: разрывности;
- 2) основанные на базовом свойстве яркости: однородности.

В первом случае изображение разбивается на области на основании некоторого изменения яркости, такого как, например, перепады яркости на изображении. Во втором случае используется разбиение изображения по критериям однородности областей. Примером первой категории может служить пороговая обработка или же пороговая классификация, а второй – выращивание областей, слияние и разбиение областей. О сегментации первого типа, а именно о пороговой обработке, и пойдет дальше речь.

Обычно пороговая сегментация изображений сводится к задаче сегментации полутоновых изображений. Действительно, выбор порога, как правило единственного, и сегментация на его основе и осуществляют переход от изображения в цветовом пространстве RGB к полутоновому, несмотря на то, что непосредственно предобработки перевода цветного изображения в полутоновое нет. Однако, иногда «цветная сегментация» все же применяется.

Предположим, что на RGB изображении необходимо выделить объекты, цвет которых лежит в определенном диапазоне. Задача сегментации в таком случае состоит в том, чтобы классифицировать каждый пиксель изображения в соответствии с тем, попадает ли его цвет в заданный диапазон или нет. Для этого в цветовом пространстве вводится мера сходства, как правило, евклидово расстояние [34]. Евклидово расстояние между точками A и B определяется выражением

$$d(A, B) = \|A - B\| = \left[(A - B)^T (A - B) \right]^{\frac{1}{2}} = \left[(A_R - B_R)^2 + (A_G - B_G)^2 + (A_B - B_B)^2 \right]^{\frac{1}{2}}, \quad (1.3)$$

где A_R, A_G, A_B – RGB компоненты вектора A , а B_R, B_G, B_B – вектора B .

В большинстве случаев сегментация изображений выполняется

на основе одного порога T , то есть осуществляется переход к полутоновым изображениям. Причиной тому является тот факт, что задачи сегментации в цветовом пространстве RGB являются узконаправленными, и для каждого изображения в таком случае необходимо знать норму расстояния для каждой компоненты R, G и B , определить которые возможно лишь путем долгих экспериментов на конкретной предметной задаче.

1.2.1. Метод порогового ограничения

Классические методы сегментации основаны на использовании порога интенсивности. Многие изображения можно охарактеризовать тем, что они содержат некоторый интересующий нас объект достаточно однородной яркости на фоне другой яркости. Типичными примерами могут служить машинописные и рукописные тексты, медицинские пробы под микроскопом, самолеты на взлетной полосе. Для таких изображений яркость служит отличительным признаком, который можно использовать для локализации объекта.

Пороговое разделение является одним из основных методов, используемых в промышленных системах технического зрения для обнаружения объектов, особенно в случаях, когда требуется наличие высокой пропускной способности системы.

Пороговая обработка является одним из основных методов сегментации изображений, благодаря интуитивно понятным свойствам. Этот метод ориентирован на обработку изображений, отдельные однородные области которых отличаются средней яркостью. Самым распространенным методом сегментации путем пороговой обработки является бинарная сегментация, то есть когда в нашем распоряжении имеется два типа однородных участков.

В этом случае изображение обрабатывается по пикселям и преобразование каждого пикселя входного изображения $f_0(x, y)$ в выходное $f_1(x, y)$ определяется из соотношения:

$$f_1(x, y) = \begin{cases} z_0, & \text{если } f_0(x, y) \leq T \\ z_1, & \text{если } f_0(x, y) > T \end{cases}, \quad (1.4)$$

где T – параметр обработки, называемый порогом, z_0 и z_1 – уровни выходной яркости. Обработка по пикселям, положение которых на изображении не играет никакой роли, называется точечной [2]. Уровни z_0 и z_1 играют роль меток. По ним определяют, к какому типу отнести данную точку: к H_0 или к H_1 . Или говорят, что H_0 состоит из фоновых точек, а H_1 из точек интереса (иногда такие точки называют

точками «кристаллизации») [1,13]. Как правило, уровни z_0 и z_1 соответствуют уровням белого и черного. Для простоты изложения далее, где это будет встречаться, будем называть классы H_1 (он же класс интереса) классом объекта, а класс H_0 классом фона.

Естественно сегментация может быть не только бинарной и в таком случае существующих классов больше, чем два. Такой вид сегментации называется многоуровневым. Результирующее изображение не является бинарным, но оно состоит из сегментов различной яркости. Формально данную операцию можно записать следующим образом:

$$f_1(x, y) = \begin{cases} 1, & \text{если } f_0(x, y) \in D_1; \\ 2, & \text{если } f_0(x, y) \in D_2; \\ \dots & \\ n, & \text{если } f_0(x, y) \in D_n; \\ 0, & \text{в остальных случаях.} \end{cases}, \quad (1.5)$$

где n – количество уровней, а $D_i, i = \overline{1, n}$ – классы изображения. В таком случае для каждого из классов должен быть задан соответствующий порог, который бы отделял эти классы между собой.

Бинарные изображения легче хранить и обрабатывать, чем изображения, в которых имеется много уровней яркости [7]. В данной работе будет рассматриваться именно сегментация, приводящая к бинарному результату.

Самым сложным в пороговой обработке является сам процесс определения порога T . Порог T часто записывают как функцию, имеющую вид:

$$T = T(x, y, l(x, y), f), \quad (1.6)$$

где f – изображение, а $l(x, y)$ – некоторая характеристика точки (x, y) изображения, например, средняя яркость в окрестности с центром в этой точке.

Если значение порога T зависит только от f , то есть одинаково для всех точек изображения, то такой порог называют глобальным. Если порог T зависит от пространственных координат (x, y) , то такой порог называется локальным. Если T зависит от характеристики $l(x, y)$, то тогда такой порог называется адаптивным. Таким образом, обработка считается глобальной, если она относится ко всему изображению в целом, а локальной, если она относится к некоторой выделенной области [19].

Помимо перечисленных разграничений алгоритмов существует еще множество методов. Многие из них являются просто совокупностью других, но большинство из них, так или иначе, базируются на анализе гистограммы исходного изображения, однако есть и принципиально другие подходы, которые не затрагивают анализ гистограмм в прямом виде или переходят от них к анализу некоторых других функций.

Прежде чем перейти к рассмотрению разных подходов к пороговой обработке изображений стоит отметить тот факт, что, вообще говоря, существует разница между рассмотрением алгоритмов в общем случае и алгоритмов к какому-то конкретному узкому кругу задач. В данной работе как раз делается упор на изучение и разработку основных алгоритмов сегментации пороговой обработкой в общем случае, сравнении и анализе их работы, поэтому предметная задача не ограничивается каким-то одним применением, но тем ни менее практическое применение тоже рассмотрено достаточно подробно. В случае конкретной задачи, модернизируя какой-либо алгоритм пороговой обработки, иногда можно достичь более хороших результатов, нежели в общем случае, так как в таком случае имеется дополнительная информация об «интересуемом» классе, но все равно, какая бы не была задача, в основе алгоритма решения лежит какой-то общий подход пороговой сегментации.

Перейдем теперь непосредственно к основным алгоритмам пороговой сегментации.

Глобальная пороговая обработка

Этот метод основан на анализе глобальной гистограммы изображения. Суть глобальной пороговой обработки состоит в разделении гистограммы изображения на две части с помощью единого глобального порога T .

Определение величины порога с помощью гистограммы яркостей является простым методом, который позволяет достичь «чистой» сегментации, если гистограмма изображения носит четко выраженный бимодальный характер [1]. Такая форма гистограммы означает, что на изображении можно различить два вида сравнительно часто встречающихся пикселей – яркие и темные. При этом гистограмма легко разделяется с помощью одиночного глобального порога T , расположенного во впадине между пиками гистограммы. Примером может служить рис.1.2.

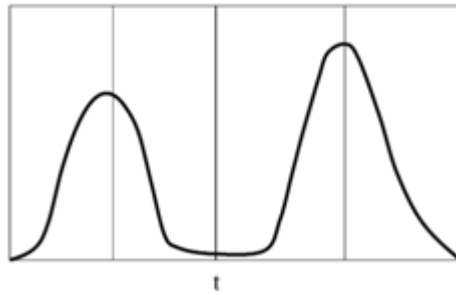


Рис. 1.2. Гистограмма яркости, допускающая сегментацию с помощью одиночного глобального порога

После этого сегментация осуществляется путем поэлементного сканирования изображения, при этом каждый пиксель отмечается как относящийся к одному классу или к другому, в зависимости от того, превышает ли яркость данного пикселя значение порога T или нет. При использовании такого порога изображение будет разделено на два класса. Так как основная цель порогового преобразования состоит лишь в получении бинарного изображения, то выбор цвета объекта или фона – черный или белый, или какой-либо другой – может быть произвольным. Естественно, чем лучше гистограмма поддается разделению, тем лучше работает алгоритм.

Порог T можно выбирать эвристически, то есть на уровне визуального изучения гистограмм. Однако в данной работе был реализован алгоритм автоматического вычисления значения порога T . Он состоит из следующих этапов, предложенных в [1]:

- 1) выбирается некоторая начальная оценка значения порога T ;
- 2) проводится сегментация изображения с помощью порога T . В результате образуются две группы пикселей: G_1 , состоящее из пикселей с яркостью больше T , и G_2 , состоящее из пикселей с яркостью меньше или равной T ;
- 3) вычисляются значения μ_1 и μ_2 средних яркостей пикселей по областям G_1 и G_2 соответственно;
- 4) вычисляется новое значение порога в соответствии с формулой:

$$T = \frac{1}{2} \cdot (\mu_1 + \mu_2) \quad (1.7)$$

- 5) повторяются шаги со 2-го по 4-й до тех пор, пока разница значений T при соседних итерациях не окажется меньше значения наперед заданного параметра ε .

Если есть основания полагать, что оба класса занимают примерно одинаковые площади на изображении, то тогда хорошим начальным приближением порога T является средний уровень

яркости изображения. Если занимаемая классом H_1 площадь мала по сравнению с площадью H_0 (или наоборот), то тогда одна из групп пикселей на гистограмме окажется больше и тогда средняя яркость окажется не самым удачным начальным приближением. В таких случаях часто начальным приближением задают полусумму максимального и минимального значения яркости на всем изображении.

Параметр ε является причиной остановки алгоритма, когда изменения на каждой итерации становятся несущественными.

Недостаток данного алгоритма заключается в том, что редко гистограмма поддается простому разделению и в большинстве реальных изображений такой подход не дает хороших результатов. Недостатком гистограммных методов в целом является то, что нет никакой гарантии в том, что пиксели, относящиеся к одной и той же моде распределения яркостей, лежат на изображении рядом и образуют связные области. Гистограмма по своей структуре не содержит характеристики о пространственном расположении пикселей [13]. Она лишь содержит информацию о частоте встречающихся на изображении пикселей с различными уровнями яркости. Однако, несмотря на этот недостаток, гистограммные методы, включая и глобальную пороговую обработку, встречаются крайне часто, потому что являются очень быстрыми методами и зачастую их результатов вполне достаточно. Если же гистограмма не подвергается четкому разделению и существует много локальных минимумов, то в таком случае сегментация глобальной пороговой обработкой оказывается не слишком успешной. Об этом можно судить по рис. 1.3 и 1.4.



Рис. 1.3. Изображение сварочного шва на листе железа



Рис. 1.4. Результат сегментации глобальной обработкой

Ниже на рис.1.5 представлена гистограмма исходного изображения сварочного шва. Видно, что форма гистограммы имеет

весьма сложную структуру и не поддается простому разделению на два класса.

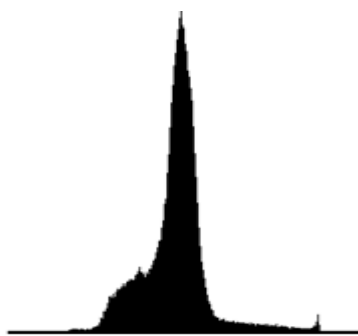


Рис. 1.5. Гистограмма изображения сварочного шва

Естественно, такие результаты не могут быть оценены как удовлетворительные, поэтому необходимо решить возникшую проблему, выбрав несколько другой подход к решению задачи. Для преодоления влияния сложной формы гистограммы на результат сегментации обратимся к локальной пороговой обработке.

Часто для улучшения метода пороговой сегментации изображения применяют инвертирование изображения. Зачастую, такая предобработка используется непосредственно в глобальной обработке. Причиной тому является тот факт, что инвертирование изображения не только инвертирует цвета, но и яркость и освещённость, что делает светлее тёмные области и темнее светлые, а это в лучшую сторону влияет на работу алгоритма глобальной пороговой обработки. Инвертирование осуществляется по следующей формуле:

$$\tilde{f}(x, y) = 255 - f(x, y), \quad (1.8)$$

где $f(x, y)$ – значение яркости пикселя с координатами (x, y) , вычисленной согласно (1.1), а $\tilde{f}(x, y)$ – новое значение яркости этого пикселя.

Однако инвертирование не всегда улучшает работу алгоритма глобальной пороговой обработки. Если гистограмма изображения представляет собой очень сложную форму, то тогда инвертирование не имеет смысла, ведь от этого общая сложность формы гистограммы не изменится, и алгоритм корректно работать не будет.

Инвертирование, очевидно, применяется не только именно как предобработка перед глобальной пороговой обработкой. Но, как показывает практика, применение этой предобработки для других методов не дает сильного улучшения работы алгоритмов.

Локальная пороговая сегментация

Одним из разновидностей методов, основанных на анализе гистограмм, является локальная пороговая обработка, но он

существенно отличается от глобального алгоритма, является более сложным и дает лучшие результаты. Данный метод является логическим продолжением глобальной пороговой обработки и применяется в основном тогда, когда гистограмма не поддается разделению с помощью единого глобального порога. Причиной такой проблемы может быть, например, неравномерное освещение на изображении.

Для решения данной проблемы изображение разбивается на некоторые подобласти, в каждой из которых используется свое значение порога. Естественно, в таком случае помимо проблемы выбора порога появляется новая проблема: как разбить исходное изображение и как на каждой области выбрать свой порог. Поскольку порог, применимый для каждого пикселя, оказывается зависящим как от характеристик подобласти изображения, которая содержит данный пиксель, так и от его пространственного нахождения, то такое пороговое преобразование является локальным. Затем все подобласти анализируются по следующим принципам:

- 1) если область не содержит границы между классами, то она характеризуется своей дисперсией яркостей d_0 ;
- 2) все области, через которые проходит граница класса N1 (и все пиксели которых содержат класс N1) задаются своей дисперсией d_1 ;
- 3) во всех областях из пункта 2) сегментация проводится с порогом, вычисленным по 3 пункту 2 данного алгоритма [1];
- 4) все области из пункта 1) обрабатываются как одно составное изображение, сегментация которого осуществлялась тоже по алгоритму из пункта 2 [1].

Таким образом, на основании 1) и 2) можно сделать вывод, что решение в выборе принадлежности пикселя к какому-либо из классов зависит от пространственного нахождения этого пикселя.

За начальное приближение во всех случаях принималась полусумма максимального и минимального значения яркости изображения.

Выбор размерности разбиения изображения полностью определяется пользователем. Однако следует заметить, что не любое разбиение приведет к верной сегментации. Единственный подход для определения верного выбора количества областей – практический. Однако общие рекомендации все-таки есть. Во-первых, задавать слишком большие разбиения (например, 50x50) не имеет смысла, иначе алгоритм становится слишком чувствительным к локальным минимумам. Естественно, не имеет смысла и разбиение 1x1. Оптимальными разбиениями являются разбиения от 2x2 до 10x10 в зависимости от размеров самого изображения: чем больше

изображение, тем допустимо более сильное разбиение.

Выбор дисперсий d_0 и d_1 также основывается исключительно из практических соображений. Для полутоновых изображений дисперсии рекомендуется выбирать в пределах $[0;50]$, а для изображений в цветовом пространстве RGB $[25;100]$. Однако бывают и исключения. Именно из-за этих трех параметров алгоритм и называется «ручным» алгоритмом. Для автоматической сегментации экспериментальным путем были установлены следующие параметры:

- количество разбиений $N=2$; - $d_0=75$; $d_1=100$.

Такой выбор параметров зачастую позволяет добиться неплохих результатов сегментации, однако, в каждом конкретном случае результат можно улучшить путем более точного подбора параметров.

Недостатком такого подхода является более долгая работа по сравнению с глобальной пороговой обработкой и невозможность автоматически выбирать точные параметры.

Достоинством локальной пороговой обработки является то, что он хорошо подходит для зашумленных изображений, для изображений, имеющих низкое качество, для изображений, характеризующихся сложной структурой фона, для изображений, яркости фона которых в различных местах достаточно сильно различаются. И, в отличие от глобальной обработки, такие алгоритмы дают неплохие результаты. Об этом можно судить по рис. 1.6, на котором изображен результат сегментации того же сварочного шва, гистограмма которого представлена на рис.1.5.



Рис. 1.7. Результат локальной пороговой обработки

Помимо рассмотренных «прямых» методов анализа изображения, есть принципиально другие подходы, позволяющие получать неплохие результаты сегментации даже ввиду сильной «изрезанности» гистограмм, не поддающихся четкому разделению, но подход которых отличается от идеи локальной пороговой обработки. Одним из таких методов является метод дискриминантного критерия.

Метод дискриминантного критерия

Пусть $w(x)$, $0 \leq x \leq 255$ – гистограмма некоторого цифрового изображения. $0 \leq x \leq 255$ – диапазон яркостей. Как уже говорилось выше, одной из проблем выбора порога является тот факт, что в

реальных ситуациях практически не существует гистограмм, поддающихся легкому разделению. Другой проблемой является еще то, что границы между односторонними участками на изображении бывают размыты, а вследствие этого уровень гистограммы в граничных точках отличается от среднего уровня. Иногда провалы между пиками классов H_1 и H_0 практически полностью исчезают [1]. Одним из методов для преодоления этих трудностей является метод дискриминантного критерия.

Пусть у нас есть некоторое изображение с двумя гипотетическими участками. Назовем их для простоты объектом и фоном. Построим распределение этого изображения $w(x)$. Для определения оптимального порога строим дискриминантную функцию $\eta(x_0^*)$, где $0 \leq x_0^* \leq 255$. Аргумент x_0^* имеет смысл пробного порога. Оно пробегает все возможные гипотетические значения порога и в результате то значение x_0^* , которое будет максимизировать функцию $\eta(x_0^*)$, и будет являться оптимальным пороговым значением x_0 . Теперь рассмотрим как построить эту дискриминантную функцию $\eta(x_0^*)$.

Пусть изначально есть некоторое гипотетическое значение порога x_0^* , которое разбивает распределение $w(x)$ на два класса H_0 и H_1 . Если точки имеют яркость равную x_0^* , то их можно отнести как к классу H_0 , так и к классу H_1 , поскольку количество таких точек достаточно мало относительно общего количества точек изображения. Тогда вероятность того, что взятая точка изображения будет относиться к классу H_0 , вычисляется по формуле (1.8):

$$p_0 = \sum_{x=0}^{x_0^*} w(x) = p(x_0^*) \quad (1.8)$$

Вероятность принадлежности пикселя к классу H_1 определяется формулой (1.9):

$$p_1 = \sum_{x=x_0^*+1}^{255} w(x) \quad (1.9)$$

Нетрудно заметить, что $p_0 + p_1 = 1$, а, значит, у нас нет точек, не принадлежащих ни H_0 , ни H_1 . Также можно заметить, что та часть распределения $w(x)$, для которой $x \in [0; x_0^*]$ (ограничена точкой x_0^*) описывает ту часть изображения, которая принадлежит классу H_0 , а

оставшаяся часть распределения $w(x)$, для которой $x \in [x_0^* + 1; 255]$, принадлежит классу H_1 . Таким образом, исходное распределение $w(x)$ можно разбить на два других распределения $w_0(x)$ и $w_1(x)$, которые будут относиться к H_0 и H_1 соответственно. Они будут описываться выражениями:

$$w_0(x) = \frac{w(x)}{P_0}, \quad 0 \leq x \leq x_0^* \quad (1.10)$$

и

$$w_1(x) = \frac{w(x)}{P_1}, \quad x_0^* + 1 \leq x \leq 255 \quad (1.11)$$

Данные выражения нормированы в силу деления на вероятности P_0 и P_1 .

Таким образом, было получено два распределения вероятностей и можно теперь найти для них моменты. Математическое ожидание – это среднее значение случайной величины. Так как в данном случае случайной величиной является x , то выражения для математических ожиданий μ_0 и μ_1 будут иметь вид:

$$\mu_0 = \sum_{x=0}^{x_0^*} x \cdot w_0(x) = \frac{\sum_{x=0}^{x_0^*} x \cdot w(x)}{P_0}, \quad (1.12)$$

$$\mu_1 = \sum_{x=x_0^*+1}^{255} x \cdot w_1(x) = \frac{\sum_{x=x_0^*+1}^{255} x \cdot w(x)}{P_1} \quad (1.13)$$

Формулу (1.13) можно преобразовать следующим образом:

$$\mu_1 = \frac{\left[\sum_{x=0}^{255} x \cdot w(x) - \sum_{x=0}^{x_0^*} x \cdot w(x) \right]}{P_1} = \frac{\left[\mu(255) - \mu(x_0^*) \right]}{P_1} = \frac{\mu(255) - \mu(x_0^*)}{1 - p(x_0^*)}, \quad (1.14)$$

где ненормированное математическое ожидание $\mu(x_0^*)$ для H_0 вычисляется как

$$\mu(x_0^*) = \sum_{x=0}^{x_0^*} x \cdot w(x), \quad (1.15)$$

а математическое ожидание для всего кадра $\mu(255)$:

$$\mu(255) = \sum_{x=0}^{255} x \cdot w(x) \quad (1.16)$$

Тогда, дисперсия для всего кадра σ_T^2 определяется выражением:

$$\sigma_T^2 = \sum_{x=0}^{255} [x - \mu(255)]^2 \cdot w(x) \quad (1.17)$$

Для того чтобы построить дискриминантную функцию, необходимо ввести еще один параметр σ_B^2 , который часто называют межклассовой дисперсией. Вычисляется он в соответствии с формулой:

$$\sigma_B^2 = p_0 \cdot (\mu_0 - \mu(255))^2 + p_1 \cdot (\mu_1 - \mu(255))^2 \quad (1.18)$$

Тогда, безразмерная дискриминантная функция строится по формуле:

$$\eta(x_0^*) = \frac{\sigma_B^2(x_0^*)}{\sigma_T^2} \quad (1.19)$$

Оптимальный порог будет являться решением экстремальной задачи вида:

$$x_0 = \arg \max_{x_0^*} \eta(x_0^*) \quad (1.20)$$

Знаменатель в соотношении (1.19) является дисперсией всего кадра, вычисляющейся в соответствии с (1.17), а, значит, непосредственно от значения порога x_0^* , который разбивает изображение на H_0 и H_1 , не зависит. Поэтому, для того, чтобы решить экстремальную задачу (1.20) достаточно решить задачу вида:

$$x_0 = \arg \max_{x_0^*} \sigma_B^2(x_0^*) \quad (1.21)$$

Теперь поясним, почему решение экстремальной задачи вида (1.20) (или же (1.21), что тоже самое) будет определять оптимальный порог.

Действительно, оптимальный порог в таком случае определяется характером зависимости межклассовой дисперсии (1.18) от пробного порога x_0^* [1]. Логично, что если x_0^* устремить к 0, то вероятность p_0 в соответствии с (1.8) также устремится к 0. А это, в свою очередь, означает, что все изображение будет целиком относиться к классу H_1 . Таким образом, математическое ожидание класса H_1 будет стремиться к математическому ожиданию всего изображения ($\mu_1 \rightarrow \mu(255)$). Значит, по формуле (1.18) межклассовая дисперсия тоже станет равной нулю. Если же пробный порог x_0^* устремить к предельному его значению 255, то мы получим

абсолютно такой же результат. Действительно, ведь тогда $p_1 = 0$, значит, все изображение принадлежит H_0 , а это в свою очередь означает, что $\mu_0 \rightarrow \mu(255)$, значит, $\sigma_B^2 = 0$.

Все величины, входящие в соотношение (1.18) являются неотрицательными. В предельных значениях $x_0^* = 0$ и $x_0^* = 255$ мы получили 0. Следовательно, на интервале $(0; 255)$ существует максимум. Значение аргумента, определяющего этот максимум, и будет оптимальным пороговым значением.

Алгоритм вычисления порогового значения x методом дискриминантного критерия можно представить следующим образом:

- 1) найти для всего изображения математическое ожидание $\mu(255)$ по формуле (1.16) и дисперсию σ_T^2 (1.17);
- 2) для всех значений x_0^* из промежутка $[0; 255]$ вычислить вероятности p_0 и p_1 в соответствии с формулами (1.8) и (1.9) соответственно;
- 3) вычислить для всех значений $x_0^* \in [0; 255]$ математические ожидания μ_0 и μ_1 по формулам (1.12) и (1.13-1.14);
- 4) составить функцию $\eta(x_0^*)$ по формуле (1.19);
- 5) решить задачу (1.20) (или же задачу (1.21)). Ту точку, при которой достигается решение этой задачи принять в качестве оптимального порога.

Однако, если подойти с практической точки зрения, то возникают некоторые трудности. Пусть изображение содержит два класса: H_0 и H_1 . Если на изображении площадь участка H_0 резко отличается от площади участка H_1 в большую или меньшую сторону, то тогда, вообще говоря, функция $\eta(x_0^*)$ может иметь несколько локальных максимумов. В таком случае необходимо определять абсолютный максимум этой функции. Правда, как правило, результаты получаются несколько хуже, чем если бы на изображении классы H_0 и H_1 имели не слишком резкую разницу в площадях.

На рис. 1.7 приведен пример аэрофотоснимка земной поверхности.

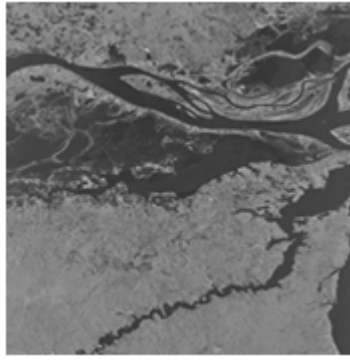


Рис.1.7. Аэрофотоснимок земной поверхности

Гистограмма данного изображения схематично представлена на рис. 1.8, а его дискриминантная функция $\eta(x_0^*)$ на рис. 1.9.

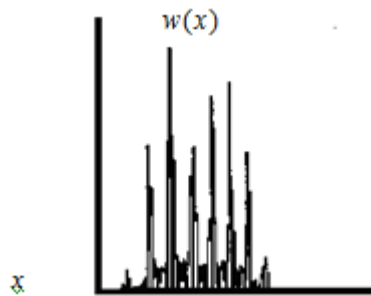


Рис.189. Гистограмма изображения 1.8.

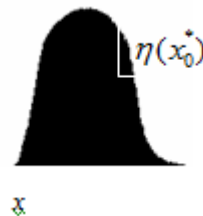


Рис. 1.9 Схематичное изображение функции $\eta(x_0^*)$

Как видно из рис.1.9, гистограмма изображения сильно изрезана, а значит невозможно найти минимум, разделяющий изображение на классы H_0 и H_1 . На рис. 1.10 представлена дискриминантная функция $\eta(x_0^*)$, построенная по алгоритму, описанному выше. Эта функция является гораздо более гладкой, нежели непосредственно гистограмма изображения, а, следовательно, максимум данной функции легко найти, а значит, тем самым сразу определяется и порог x .

На рис. 1.10 представлен результат сегментации данного изображения, вычисленный по вышеизложенному алгоритму.



Рис. 1.10. Результат сегментации аэрофотоснимка

Из представленных рисунков видно, что алгоритм корректен и по результирующему изображению вполне можно получить представления о характере однородных областей. Сильная изрезанность гистограммы изображения говорит о невозможности применения общих методов, основанных на «чистом» разделении гистограммы, поэтому данный метод в таких условиях решает проблему.

Данный алгоритм может применяться именно для обработки аэрофотоснимков изображения, однако этим применение данного алгоритма не ограничивается. При анализе аэрофотоснимков местностей пороговую обработку методом дискриминантного критерия используют в совокупности с другими методами. Например, для определения каких-либо объектов на местности, скажем ферм, используют вместе с пороговой обработкой методы, позволяющие учитывать геометрические формы областей, например, размеры. И пороговая обработка становится неотъемлемой частью обработки изображения. Касательно других изображений метод дискриминантного критерия можно применять в большинстве случаев, однако, как уже было сказано, если площади, покрываемые классами H_0 и H_1 , очень сильно различаются, то пороговая обработка методом дискриминантного критерия не позволит добиться наилучших результатов, если же этого нет, а гистограмма не поддается четкому распределению, то применение данного алгоритма оправданно.

Адаптивная пороговая сегментация

В соответствии с (1.6), если порог T зависит от характеристик подобласти, то такая обработка называется адаптивной. Естественно встает вопрос о том, какую же зависимость представляет собой порог T от этих характеристик.

Как уже рассматривалось выше в локальной пороговой обработке, существуют сильно зашумленные изображения, которые невозможно обработать как изображение в совокупности, поэтому его разбивают на подобласти. Однако в локальной пороговой обработке решение о принадлежности пикселя какому-либо из

классов принималось на анализе данных о пространственном положении этого пикселя. Но было напрямую неважно, какие по соседству с ним пиксели находятся, и какой характеристикой они обладают.

Итак, адаптивный алгоритм пороговой обработки базируется на идее сопоставления уровней яркости преобразуемого пикселя со значениями локальных средних, вычисленных непосредственно в его окружении, то есть у соседей. Пиксели обрабатываются поочередно. Интенсивность каждого пикселя сравнивается со средними значениями яркости в окнах размерности $(2D+1) \times (2D+1)$ с центром в точке K . Более подробно схема изображена на рис.1.11.

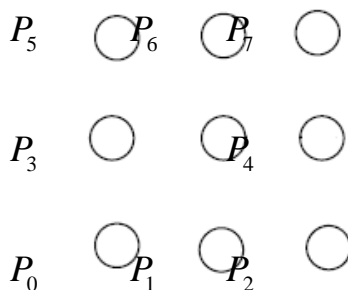


Рис.1.11. Направление обхода соседних пикселей

Пусть пиксель K с центром в данной области имеет координаты $K(i, j)$. Тогда K считается пикселем класса H_0 , если у него стоит метка 0, и пикселем H_1 , если у него стоит метка 1. То есть K становится пикселем H_1 тогда, когда для $r = \overline{0, s}$, где $s = (2D+1) \cdot (2D+1) - 1$ выполняется условие:

$$Z_r - f(i, j) > T, \quad (1.22)$$

где T и играет как раз роль порогового значения, а Z_r средняя локальная яркость, определяющаяся по формуле:

$$Z_r = \frac{1}{(2D+1) \cdot (2D+1)} \cdot \sum_{-D \leq m, n \leq D} f(i_r + m, j_r + n), \quad (1.23)$$

где $f(i_r, j_r)$ – яркость в точке K_r с координатами (i_r, j_r) .

Параметр D теоретически может быть любым, однако, целесообразно его выбирать в пределах от 0 до 10, поэтому в программе была предоставлена возможность выбирать именно в таком диапазоне. Такой диапазон обусловлен, тем, что при сильном увеличении этого параметра, то есть области соседей, практически было замечено, что результат почти не отличается от установки параметра $D=10$, а время работы алгоритма становится

несоизмеримо больше. Более того, при работе с большими изображениями, если параметр D задавать слишком большим, требуется очень много виртуальной памяти, и на слабых машинах программа будет «вылетать». Параметр D обычно подбирают, однако есть общие рекомендации. Например, при анализе и распознавании документов этот параметр выбирают как среднюю толщину линий рисунков и символов. При сегментации бактерий выбирают среднюю толщину «брюшка» бактерии и так далее. Часто параметр выбирают исходя из непосредственных практических экспериментов.

Для того чтобы устранить ошибки обычного порогового преобразования, очевидно, нельзя применять глобальный алгоритм. Поэтому применяют следующее автоматическое и адаптивное определение параметра T :

1) пусть $f(i, j)$ - пиксель изображения с координатами (i, j) . В окне $(2D+1) \times (2D+1)$ с центром в $f(i, j)$ вычисляются максимальные и минимальные значения яркости:

$$f_{\max} = \max_{-D \leq a, b \leq D} f(i+a, j+b), \quad (1.24)$$

и

$$f_{\min} = \min_{-D \leq a, b \leq D} f(i+a, j+b). \quad (1.25)$$

2) вычисляются максимальное и минимальное приращение яркости относительно центрального пикселя по формулам:

$$\Delta f_{\max} = |f_{\max} - \tilde{Z}|, \quad (1.26)$$

$$\Delta f_{\min} = |f_{\min} - \tilde{Z}|. \quad (1.27)$$

3) далее сравниваются величины Δf_{\max} и Δf_{\min} . Если $\Delta f_{\max} > \Delta f_{\min}$, то тогда окно $(2D+1) \times (2D+1)$ содержит больше локальных низких яркостей. В соответствии с этим выбирается T по формуле:

$$t = \alpha \left(\frac{2}{3} f_{\min} + \frac{1}{3} \tilde{Z} \right), \quad (1.28)$$

где $\alpha \in (0; 1]$ - некоторая константа.

4) если же $\Delta f_{\max} < \Delta f_{\min}$, то тогда в окне $(2D+1) \times (2D+1)$ содержится больше локальных высоких яркостей. И в соответствии с этим порог T вычисляется по формуле:

$$t = \alpha \left(\frac{1}{3} f_{\min} + \frac{2}{3} \tilde{Z} \right). \quad (1.29)$$

Параметр $\alpha \in (0;1]$ является параметром регуляризации. Как правило, для большинства изображений этот параметр выбирают равным $\frac{2}{3}$. Но если изображение имеет очень высокий шум и низкий

контраст, то тогда $\alpha = \frac{1}{3}$.

В большинстве случаев параметры задают, как правило, так:

$$D=1, \quad \alpha = \frac{2}{3}$$

Данный алгоритм позволяет выполнить сегментацию для изображений, содержащих шум, зависящих от чувствительного сигнала, изображений, имеющих сложную структуру фона или же малый контраст, и при этом практически без потерь полезной информации. Примером работы алгоритма с такими изображениями является рис. 1.12 - 1.14.

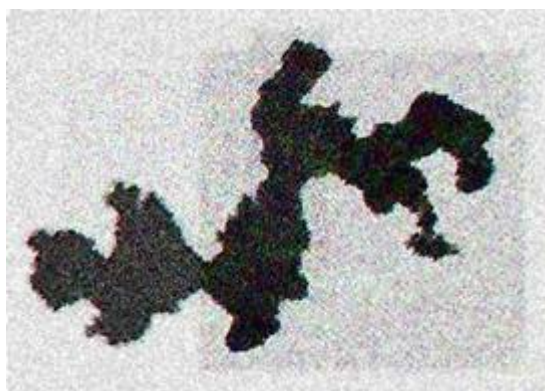


Рис. 1.12 Искусственно сильно зашумленное изображение

Гистограмма изображения представлена на рис. 1.14. Очень сильная изрезанность говорит о сильном влиянии шумов и неоднородности фона на изображении.



Рис. 1.13 Гистограмма изображения на рис.1.12.



Рис. 1.14 Результат адаптивной обработки ($D=5$, $\alpha=0.62$)

Однако у данного алгоритма, как и у любого другого, есть недостатки. С ростом значения параметра D существенно возрастает время работы алгоритма и затраты вычислительных ресурсов компьютера, что особенно заметно при работе с большими изображениями. Также данный алгоритм не является полностью автоматическим, а, следовательно, усложняет работу пользователя с ним по крайней мере в том случае, если пользователь хочет добиться наилучших результатов. Алгоритм нецелесообразно применять к относительно простым изображениям, с которыми глобальный алгоритм и алгоритм на основе построения дискриминантной функции справляются лучше и быстрее.

Следует отметить тот факт, что уровень чувствительности алгоритма можно менять. Например, если задача заключается в поиске темных объектов на светлом фоне, то при повышении параметра α чувствительность к изменению яркости падает, а при понижении – растет.

То есть при повышении параметра степень детализации уменьшается, а при уменьшении – увеличивается. Это следует из формул определения пороговой величины (1.28) и (1.29). Если задача поставлена в противоположном аспекте, то есть поиске светлых участков на более темных, то ситуация складывается обратная. Продемонстрируем это. Простым схематическим примером может служить изображение на рис. 1.15. На этом рисунке изображен негатив солнца с имеющимися на солнце пятнами.



Рис. 1.15. Изображение солнца



Рис. 1.16 Результат обработки при $\alpha = 0.36$



Рис. 1.17 Результат обработки при $\alpha = 0.1$ ($D = 10$)

Анализируя изображения на рис.1.16 и 1.17, можно сказать, что результаты обработки рисунков различны. Если под задачей подразумевается выделение звезды, как таковой, то результатом может быть рис. 1.18, а если под задачей подразумевается определение пятен (выбросов) на солнце, то очевидно, что результатом будет рис. 1.16. В этом случае степень детализации на рис. 1.16 больше, чем на рис. 1.7. Параметр D определялся в соответствии с его смысловой нагрузкой: если в первом случае нас интересовали пятна, то $D=1$, так как объекты небольшие, а во втором – максимальному значению $D=10$, так как нас интересует непосредственно солнце, то есть большой объект. На данном примере очень точно еще отображается тот факт, что понятие объекта и фона – относительное. Нельзя говорить, что все время белые области на изображении будут представлять для нас интерес. Как мы видим, в одном случае белые пиксели образуют класс интереса, а в другом – черные.

1.2.2.Операции над бинарными изображениями.

Над бинарными изображениями могут быть выполнены различные логические операции:

- операции булевой алгебры

$$C = A \& B; \quad C = A \cup B; \quad C = \bar{A}$$

(заметим, что инверсия соответствует получению негатива бинарного изображения)

- операции алгебры де Моргана

$$\overline{A \cup B} = \overline{A} \& \overline{B}$$

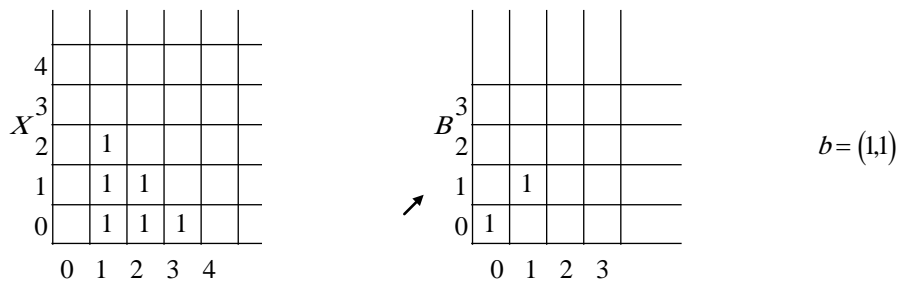
$$\overline{A \& B} = \overline{A} \cup \overline{B}$$

Кроме того над бинарным изображением могут быть выполнены так называемые морфологические операции [13]. Особенностью подобных операций является то, что они выполняются над исходным бинарным изображением и его копией, сдвинутой по отношению к исходному на целое число пикселей по координатам i и j . Эти сдвиги определяют вектор смещения. К такого типа операциям относятся операции эрозии и дилатации.

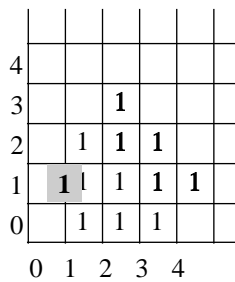
Операция эрозии:

$$E(x, b) = X \cap \hat{X}(b)$$

$$\hat{X}(b) \rightarrow \hat{x}_{i+b, j+b} = x_{i, j}$$

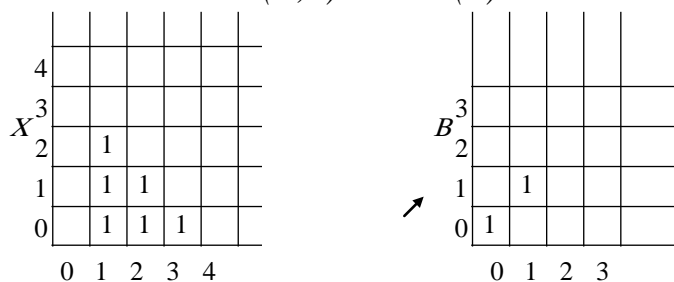


Результат операции эрозии:



Операция дилатации

$$D(X, b) = X \cup \hat{X}(b)$$



4					
3			1		
2		1	1	1	
1		1	1	1	1
0		1	1	1	
	0	1	2	3	4

Морфологические операции эрозии и дилатации позволяют заполнить пустоты или, наоборот, устранить эффект бахромы на краях объектов.

1.2.3. Метод наращивания областей

Гистограмма изображения является его глобальной характеристикой - при ее формировании не используется понятие близости элементов изображения. Поэтому методы, рассмотренные выше, являются глобальными. Существуют, однако, и локальные методы, основанные на наращивании областей.

Алгоритмы наращивания областей используют информацию о связности объектов и основаны на рекуррентном способе разметки точек. На шаге k размечаются те и только те точки, которые имеют соседей из числа размеченных на предыдущем шаге $(k-1)$. Разметка точек осуществляется согласно некоторому критерию однородности. Конкретные алгоритмы различаются выбором критерия однородности, способом просмотра точек и выбором начальных "стартовых" точек, размечаемых на нулевом шаге.

Известно два основных подхода к стратегии выбора стартовых точек и порядка просмотра остальных: центроидное связывание и слияние-расщепление объекта [13, 46].

Центроидное связывание. Выбор стартовых точек и их меток в алгоритмах центроидного связывания должен быть осуществлен так, чтобы никакие две точки с различными метками не были соседними. Кроме того, если априорно известна некоторая информация о расположении объектов в поле зрения, то желательно стремиться к выполнению следующих требований:

- точки с различными метками должны соответствовать областям различных объектов;
- точки с одинаковой меткой должны соответствовать одному и тому же объекту.

Такие требования будут выполнены в том случае, если каждой стартовой точке сопоставляется своя метка, а сами точки выбираются

на достаточно большом расстоянии друг от друга, превышающем максимальный размер объекта. При выборе нескольких точек с одной меткой они должны образовывать достаточно однородное по яркости множество; обычно оно является связным.

Таким образом, в алгоритмах центроидного связывания априорная информация об объектах учитывается в основном на этапе выбора стартовых точек.

Выбор стартовых точек - нулевой шаг алгоритмов центроидного связывания. Переход от шага с номером k к шагу с номером $(k+1)$ осуществляется следующим образом. Пусть S_k - множество точек, размеченных после реализации шага с номером k . На очередном, $(k+1)$ -м, шаге разметке подлежат точки, примыкающие к S_k , т.е. не принадлежащие S_k но имеющие хотя бы одного соседа из S_k . Пусть A - одна из таких точек. Через $\lambda_1, \dots, \lambda_p$ обозначим метки, используемые для разметки точек из S_k . Рассмотрим два возможных случая.

Все размеченные соседи A имеют одну и ту же метку λ_j . В этом случае точка A размечается либо меткой λ_j , либо меткой, т.е. еще не использованной. Первый вариант осуществляется при выполнении некоторого условия однородности. Обычно оно имеет вид $|B(A) - B_{\lambda_j}| < T$, где $B(A)$ - яркость в точке A ; B_{λ_j} - средняя яркость точек с меткой λ_j ; T - выбранный порог.

Соседние A точек из S_k имеют различные метки $\lambda_1, \dots, \lambda_l$. В этом случае проверяется условие $|B(A) - B_{\lambda_i}| < T$ при каждом $i=1, \dots, l$. Если оно выполнено для единственного $i=i^0$, то точка A получает метку λ_{i^0} . Если же условие не выполнено ни при каком $i=1, \dots, l$, то точка A получает новую метку λ_{p+1} . В наиболее сложном случае условие справедливо для нескольких значений, например для $i=1$ и $i=2$. В этом случае сливаются две области - области точек с метками λ_1 и λ_2 соответственно. Под слиянием понимается переход от меток λ_1, λ_2 к единой метке $\min(\lambda_1, \lambda_2)$.

В случае слияния областей меняется частичная разметка, полученная на шаге k . Поэтому анализ большинства точек изображения следует провести заново. Конечно, необходимость повторных проходов по полю изображения резко увеличивает время реализации алгоритма. Слияние областей - наиболее ответственные моменты алгоритмов центроидного связывания. Чем меньше число раз возникает вопрос о слиянии областей, тем выше надежность алгоритма, тем меньше время, необходимое для его реализации. Число слияний областей зависит от правильности выбора порога T и стартовых точек.

Результаты сегментации в алгоритмах центроидного связывания зависят и от порядка просмотра точек. Технически

наиболее просто осуществляется сканирование поля зрения по строкам. Однако для сложных изображений более надежным является "волновой" способ просмотра точек. Для каждой метки $\lambda_1, \dots, \lambda_s$, соответствующей некоторой стартовой точке, строится "волна". Волной называется набор множеств точек, называемых фронтами (они обозначаются индексами $0, 1, 2, \dots$). Нулевой фронт $F_0(\lambda_i)$, соответствующий метке λ_i , есть просто подмножество стартовых точек с этой меткой. Фронт $F_{k+1}(\lambda_i)$ определяется как множество точек, которые не принадлежат $F_k(\lambda_i)$, но имеют соседей из $F_k(\lambda_i)$.

Порядок просмотра точек определим теперь следующим образом. Сначала обрабатываются точки первого фронта всех волн, соответствующих меткам $\lambda_1, \dots, \lambda_s$, затем - точки второго фронта и т.д. Именно такой порядок просмотра точек отличается тем, что при нем число обрабатываемых точек фона минимально. В задачах робототехники этот фактор существенен, так как вопрос о слиянии областей обычно возникает при разметке какой-либо точки области фона.

Основными недостатками метода центроидного связывания являются:

- отсутствие правила для выбора порога, фигурирующего в критерии однородности области;

- необходимость повторных проходов в случае слияния областей.

Тем не менее, алгоритмы центроидного связывания успешно используются в задачах робототехники, особенно в сочетании с алгоритмами порогового ограничения. Они обладают важнейшими достоинствами - простотой реализации и возможностью оптимального учета дополнительной априорной информации о расположении объектов.

Алгоритмы слияния-расщепления. Алгоритмы слияния-расщепления отличаются от алгоритмов центроидного связывания главным образом порядком просмотра точек. Они направлены на выделение в поле зрения D однородных областей как можно больших размеров. Предварительно выделенные однородные области интерпретируются как один элемент изображения.

Будем считать, что поле зрения D является растром размером $m \times m$, причем $m = 2^k$. Отдельные точки растра (пиксели) назовем областями нулевого уровня и обозначим их S^0_{ij} , $i, j = 1, \dots, 2^k$. Определим теперь области первого уровня как объединения четырех смежных областей нулевого уровня, а именно:

$$S^1_{i,j} = S^0_{2i-1, 2j-1} \cup S^0_{2i-1, 2j} \cup S^0_{2i, 2j-1} \cup S^0_{2i, 2j},$$

причем $i, j = 1, 2, \dots, 2^{k-1}$.

При произвольном $p \leq k$ определим теперь области p -го уровня

в виде объединения областей $(p-1)$ -го уровня:

$$S^p_{i,j} = S^{p-1}_{2i-1,2j-1} \cup S^{p-1}_{2i-1,2j} \cup S^{p-1}_{2i,2j-1} \cup S^{p-1}_{2i,2j},$$

$i, j = 1, 2, \dots, 2^{k-p}$. Последний уровень имеет номер k . Для k -го уровня есть только одна область, совпадающая с растром D .

Отметим основные свойства указанных разбиений:

- число областей уровня p , $0 \leq p \leq k$, равно $2^{2(k-p)}$;
- объединение областей любого уровня есть растр D ;
- каждая область уровня p состоит из 2^{2p} точек растра;
- каждая область уровня $(p+1)$ есть объединение четырех смежных областей уровня p .

Первый шаг алгоритма состоит в анализе всех областей 1-го уровня и выделения среди них однородных областей, разброс яркости точек которых не превосходит фиксированного порога. Однородные области при дальнейшем анализе следующих уровней интерпретируются как один элемент изображения с яркостью, равной средней яркости точек соответствующей области. Таким образом, при выполнении критерия однородности для области $S^l_{i,j}$ области $S^0_{2i-1,2j-1}, \dots, S^0_{2i,2j}$ отдельно не рассматриваются - происходит их слияние.

На следующем шаге рассматриваются области 2-го уровня, но не все, а только такие области $S^2_{i,j}$, для которых соблюдается условие: для каждой из областей $S^1_{2i-1,2j-1}, \dots, S^1_{2i,2j}$ выполнен критерий однородности (это проверяется на предыдущем шаге), т.е. они уже представляют собой один элемент изображения. Теперь критерий однородности проверяется для набора $S^1_{2i-1,2j-1}, \dots, S^1_{2i,2j}$, т.е. вычисляется разброс четырех значений яркости, соответствующих этим областям (теперь уже - элементам изображения). При выполнении условия однородности указанные области сливаются в единый элемент изображения с усредненной яркостью.

Такая процедура осуществляется последовательно на 1-, 2-, ..., l -м уровнях. Заканчивается она на некотором уровне l , когда выполнено следующее условие: для любой области $S^l_{i,j}$, либо ни одна из областей $S^{l-1}_{2i-1,2j-1}, \dots, S^{l-1}_{2i,2j}$ не является единым элементом изображения (информация об этом имеется после анализа $(l-1)$ -го уровня), либо те из этих областей, которые являются едиными элементами изображения, не удовлетворяют условию однородности, т.е. не могут быть слиты.

В результате получается разбиение растра D на определенное число однородных областей, принадлежащих различным уровням и состоящих из различного числа точек. Таким образом, тем областям поля зрения, в которых яркость колеблется в незначительных пределах, отвечают большие области в разбиении растра, неоднородные по яркости области "дробятся" на мелкие части.

Трудность реализации алгоритмов слияния-расщепления

заключается, в первую очередь, в необходимости большого объема памяти ЭВМ для хранения промежуточных результатов (пирамидальных структур). Такие алгоритмы следует использовать для сегментации в тех случаях, когда в поле зрения необходимо выделить объекты существенно различных размеров. Такая ситуация в робототехнике встречается редко, поэтому метод слияния-расщепления используется чаще для получения промежуточного результата - выделения максимально большой области, соответствующей фону. Информация о такой области позволяет во многих случаях найти порог яркости, разделяющей фон и объекты.

Методы сегментации цветных изображений

На протяжении последних десятилетий сегментация изображений, т.е. идентификация однородных областей на изображении, является объектом активных исследований. Существует большое количество алгоритмов автоматической сегментации полутоновых изображений. Однако, проблеме сегментации цветных изображений, содержащих большее количество информации об объектах сцены, уделено значительно меньшее внимание. В настоящее время с развитием и распространением аппаратуры регистрации цифровых цветных изображений разработчикам компьютерных систем обработки изображений (систем технического зрения) все чаще приходится решать задачи сегментации цветных изображений [34].

Сегментация цветного изображения это процесс выделения из пространства изображения одной или нескольких связных областей, удовлетворяющих критерию однородности, основанному на признаках, вычисляемых из значений нескольких цветовых компонентов. Эти компоненты определены в выбранной модели цветового пространства [13, 34].

Изображения, полученные с помощью цифровых камер и сканеров, обычно представлены в цветовом пространстве RGB. Также широко используются цветовые модели, основанные на восприятии человека, такие как HSB, и аппаратно-независимые модели XYZ, CIE(Lab), CIE(Luv) и др.

Например, цветное изображение размером $M \times N$, заданное в системе координат RGB представляется в виде трех векторов, длиной каждого $M \times N$. Для дальнейших исследований вычисляются ковариационная матрица K изображения, ее собственные значения $e_{ij}, i, j = 1, 2, 3$ и определяются оси новой системы координат.

$$\begin{vmatrix} V_1 \\ V_2 \\ V_3 \end{vmatrix} = \begin{vmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{vmatrix} \begin{vmatrix} R \\ G \\ B \end{vmatrix}$$

где V_1, V_2, V_3 — новые оси, e_{ij} — координаты собственных векторов, соответствующих собственным значениям l_1, l_2, l_3 , где $l_1 > l_2 > l_3$.

Возможно, наиболее важный признак метода сегментации это способ определения области. Можно выделить 4 основных типа определений:

1. Область это связное подмножество множества пикселей, определенное порождающей функцией, заданной в пространстве цветов. Семейство алгоритмов, использующих такое определение области, называют пиксельными алгоритмами сегментации. Выделяется несколько групп пиксельных алгоритмов:

а). Пороговое отсечение. В данной группе алгоритмов строятся гистограммы цветовых признаков.

б). Кластеризация в пространстве цветов. Значения пикселей собираются в группы (кластеры) с одним или более представителем, которые используются далее в процессе классификации.

с). Кластеризация в пространстве цветов с использованием аппарата нечеткой логики (fuzzy logic). Для всех пикселей вычисляются нечеткие функции принадлежности и определяются нечеткие кластеры. Четкие кластеры получаются в результате анализа степени принадлежности. Далее изображение разбивается на связные области.

2. Область это максимальное по площади связное множество пикселей, для которого выполняется свойство однородности. Алгоритмы, использующие такой подход, можно разделить на две группы:

а) Методы наращивания областей.

б) Методы разбиения-слияния. 3. Область это связное множество пикселей, окруженное граничными пикселями, образующими цветовой контур. Области также однородны, т.к. они являются дополнением множества неоднородных областей, образованных граничными пикселями. Часто контур имеет разрывы и необходимо решать задачу замыкания контуров.

Локальные методы для определения граничных точек используют информацию из окрестности этой точки. Глобальные методы выполняют глобальную оптимизацию, и, таким образом, граничная точка может быть определена в результате многих шагов оптимизации, включающих изменения больших участков изображения.

4. Области соответствуют поверхностям или объектам из

однородного материала, т.е. области представляют связный материал в сцене. Основное отличие данных методов от вышеперечисленных состоит в том, что учитываются отражающие свойства материалов сцены. Таким образом, применение данного класса алгоритмов ограничивается набором объектов, для которых хорошо известны и могут быть смоделированы их отражающие свойства.

1.3. Методы выделения границ объектов

Существует принципиально иной (по сравнению с методом разметки точек) способ сегментации - сегментация путем выделения границ объектов. При таком способе сегментации объекты представляются их границами. Выделение границ объектов можно рассматривать и как самостоятельную практическую задачу, не связанную непосредственно с сегментацией. Границы - основа формирования различных признаков и грамматик при распознавании изображений. Существует достаточное количество методов выделения границ и анализа контуров, однако не все из них находят применение в робототехнике. Ограничивающим критерием является объем вычислений.

Понятие границы объекта так же, как и понятие самого объекта, невозможно точно формализовать в терминах цифрового изображения $B(i,j)$. Из эвристических соображений граничные точки ищут как точки резкого перепада функции яркости.

Если представить изображение раstra D как функцию $B(x,y)$ в плоскости (x,y) , и считать, что $B(x,y)$ - дифференцируемая функция своих аргументов, то перепад яркости в точке (x_0,y_0) определяется как норма градиента функции $B(x,y)$ в точке (x_0,y_0) , т.е. норма вектора

$$\|\nabla B(x_0,y_0)\| = \left\| \begin{matrix} \frac{dB(x,y)}{dx} \\ \frac{dB(x,y)}{dy} \end{matrix} \right\|_{x_0,y_0}$$

норму вектора можно определить различными способами.

Однако в силу дискретности поля зрения D непосредственное вычисление $\nabla B(x_0,y_0)$ невозможно. Чтобы воспользоваться дифференцированием для определения перепада яркости в точке, необходимо выбрать один из двух путей [6]:

- с помощью интерполяции перейти от функции $B(i,j)$ к функции $B(x,y)$, заданной в прямоугольной области плоскости (x,y) , содержащей все точки раstra D ;
- воспользоваться методами численного дифференцирования, т.е. дифференцирования таблично заданной функции.

Первый путь получил название метода функциональной аппроксимации. Он связан с большими вычислительными затратами и в задачах робототехники не получил большого распространения.

Наиболее распространенным является второй путь, называемый градиентным методом, методом контрастирования или методом пространственного дифференцирования.

1.3.1. Градиентный метод

Градиент изображения $f(x,y)$ в точке (x,y) определяется как двумерный вектор,

$$G[f(x, y)] = \begin{vmatrix} G_x \\ G_y \end{vmatrix} = \begin{vmatrix} \frac{df}{dx} \\ \frac{df}{dy} \end{vmatrix}$$

который указывает направление максимального изменения функции f в точке (x,y) :. Особый интерес эта величина представляет при определении кромок (контура) некоторого объекта, наблюдаемого на произвольном фоне.

Для цифрового изображения это можно сделать несколькими путями. Один из подходов состоит в использовании разности между соседними пикселями:

$$G_x = \frac{df}{dx} = f(x, y) - f(x - 1, y)$$

$$G_y = \frac{df}{dy} = f(x, y) - f(x, y - 1)$$

Несколько более сложный способ включает пиксели в окрестности размерностью 3×3 с центром в точке (x,y) .

Существует множество способов формирования выходного изображения $g(x,y)$, основанных на вычислении градиента. Простейшим способом является задание функции g в точке (x,y) значения, равного величине градиента входного изображения f в этой точке.

Другим способом получения дискретного изображения является применение следующих соотношений:

$$g(x,y) = G[f(x,y)]$$

$$g(x, y) = \begin{cases} 1, & G[f(x, y)] > T \\ 0, & G[f(x, y)] \leq T \end{cases}$$

где T – неотрицательная пороговая величина. В этом случае имеют значение только пиксели кромки, градиенты которых превышают величину T . Таким образом, использование последнего

выражения может рассматриваться как процесс выделения только тех пикселей, которые характеризуются значительным (определенным величиной T) перепадом интенсивности.

1.3.2. Выделение контуров

Градиентные методы определяют разрывы в интенсивности представления образа объекта. В идеальном случае эти методы определяют пиксели, лежащие на границе между объектом и фоном. На практике данный ряд пикселей редко полностью характеризует границу из-за шума и разрывов на границе. Таким образом, алгоритмы обнаружения контуров сопровождаются процедурами построения границ объектов из соответствующих последовательностей пикселей. Ниже рассмотрено несколько методов, пригодных для этой цели.

Локальный анализ. Одним из наиболее простых подходов соединения точек контура является анализ характеристик пикселей в небольшой окрестности (например, в окрестности размером 3×3) каждой точки (x, y) образа, который уже подвергся процедуре обнаружения контура. Все точки, являющиеся подобными, соединяются, образуя границу из пикселей, обладающих некоторыми свойствами.

При таком анализе для установления подобия пикселей контура необходимо определить [13]:

1. Величину градиента, требуемого для построения контурного пикселя $G[f(x, y)]$;
2. Направление градиента.

Таким образом, пиксел контура с координатами (x', y') подобен по величине в определенной ранее окрестности (x, y) пикселу с координатами (x, y) , если справедливо неравенство

$$\left| G[f(x, y)] - G[f(x', y')] \right| \leq T \quad ,$$

где T - пороговое значение.

Направление градиента устанавливается по углу вектора градиента

$$\theta = \arctg \left[\frac{G_y}{G_x} \right] \quad ,$$

где θ – угол (относительно оси x), вдоль которого скорость изменения имеет наибольшее значение. Тогда можно сказать, что угол пиксела контура с координатами (x', y') в некоторой окрестности (x, y) подобен

углу пиксела с координатами (x,y) при выполнении следующего неравенства

$$|\theta - \theta'| < A ,$$

где A – пороговое значение угла.

Основываясь на этих предположениях, мы соединяем точку в некоторой окрестности (x,y) с пикселом, имеющим координаты (x,y) , если удовлетворяются критерии по величине и направлению. Двигаясь от пиксела к пикселу и представляя каждую присоединяемую точку как центр окрестности, процесс повторяется для каждой точки образа.

1.3.3. Градиентная обработка при линейной фильтрации

Для выделения контуров объектов на изображении может быть использован метод линейной фильтрации, основанный на вычислении апериодической свертки фрагмента изображения со специальным ядром в пространственной области. Соответственно, используемое для вычислений ядро будет определять тип линейного фильтра. Для выделения контуров методом линейной фильтрации наиболее часто используют следующие виды фильтров [13]:

1. разностный амплитудный фильтр

$$y_{ij} = \frac{1}{2} \sqrt{[(a_{i,j} - a_{i,j+1}) + (a_{i+1,j} - a_{i+1,j+1})]^2 + [(a_{i,j} - a_{i+1,j}) + (a_{i,j+1} - a_{i+1,j+1})]^2}$$

2. максимальный разностный амплитудный фильтр

$$y_{ij} = \left| \max(a_{i,j}; a_{i,j+1}; a_{i+1,j}; a_{i+1,j+1}) - \min(a_{i,j}; a_{i,j+1}; a_{i+1,j}; a_{i+1,j+1}) \right|$$

3. фильтр Робертса

$$y_{ij} = \sqrt{(a_{i,j} - a_{i+1,j+1})^2 + (a_{i+1,j} - a_{i,j+1})^2}$$

4. фильтр Превитта

$$y_{ij} = \frac{1}{2} \sqrt{[(a_{i,j} - a_{i,j+2}) + (a_{i+1,j} - a_{i+1,j+2}) + (a_{i+2,j} - a_{i+2,j+2})]^2 + [(a_{i,j} - a_{i+2,j}) + (a_{i,j+1} - a_{i+2,j+1}) + (a_{i,j+2} - a_{i+2,j+2})]^2}$$

5. фильтр Собеля

$$y_{ij} = \sqrt{[(a_{i,j} - a_{i,j+2}) + 2 \cdot (a_{i+1,j} - a_{i+1,j+2}) + (a_{i+2,j} - a_{i+2,j+2})]^2 + [(a_{i,j} - a_{i+2,j}) + 2 \cdot (a_{i,j+1} - a_{i+2,j+1}) + (a_{i,j+2} - a_{i+2,j+2})]^2}$$

Градиентные фильтры могут быть также записаны в матричной форме. Для окна размером 3×3 маски таких фильтров имеют следующий вид для различных фильтров:

а) Робертса:

Для оси ОХ

0	0	0
0	-1	0
0	0	1

Для оси ОУ

0	0	0
0	0	-1
0	1	0

б) Собеля:

Для оси ОУ

-1	-2	-1
0	0	0
1	2	1

Для оси ОХ

-1	0	1
-2	0	2
-1	0	1

в) Превитта:

Для оси ОХ

-1	-1	-1
0	0	0
1	1	1

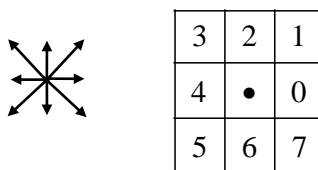
Для оси ОУ

-1	0	1
-1	0	1
-1	0	1

1.4. Описание контура объекта

На практике описание выделенного контура достаточно просто производится при использовании цепного кода Фримана [15]. Такой подход позволяет от двумерных объектов перейти к их одномерному (векторному) описанию, т.е. построение цепного кода может рассматриваться и как процедура векторизации изображения.

При выполнении кодирования цепным кодом строится так называемая матрица направлений или матрица связности центральной точки локальной окрестности размера 3×3 со всеми возможными соседними точками:



Каждое из 8-ми возможных направлений перехода от точки B_{ij} к точке $B_{i+1, j+1}$ кодируется трехбитным кодом.

Затем задается начальная точка, с которой начинают обход выделенного контура - пусть это будет крайняя левая верхняя точка (рис.1.18).



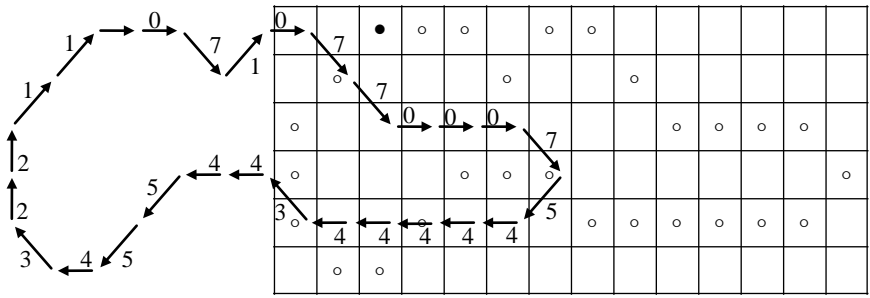


Рис.1.18. Описание контура объекта цепным кодом Фримена

В нашем случае это точка с координатами

$$A_0(1,3) \quad \begin{matrix} x = 1 ; y = 3 \\ (i = 1) (j = 3) \end{matrix}$$

Тогда контур может быть описан как:

$$A_0(1,3) \rightarrow \{0;0;7;1;0;7;7;0;0;0;7;5;4;4;4;4;4;3;4;4;5;5;4;3;2;2;1;1\}$$

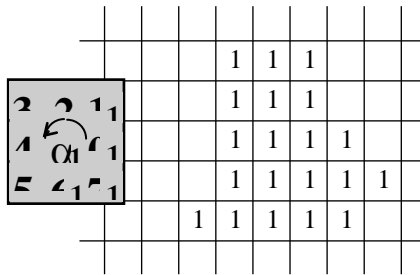
1.5. Скелетизация объекта

Под скелетизацией или утоньшением будем понимать преобразование изображения, удовлетворяющее следующим условиям:

- после преобразования все линии имеют толщину в 1 элемент;
- преобразование не нарушает топологию символа;
- преобразование не нарушает размеров символа.

Рассмотрим алгоритм скелетизации, учитывающий пошаговый режим просмотра изображения [15].

Вначале определим признаки характерных точек локальной окрестности (окна) при скелетизации объекта.



Признак того, что элемент α является верхним в окне:

$$\varphi_{\text{В}} = \bar{2} \& \alpha \& \bar{6}$$

Признак того, что элемент является левым в окне:

$$\varphi_{\text{Л}} = \bar{4} \& \alpha \& 0$$

Признак того, что элемент является правым в окне:

$$\varphi_{\text{П}} = 4 \& \alpha \& \bar{0}$$

Признак того, что элемент является нижним в окне:

$$\varphi_{\text{В}} = 2 \& \alpha \& \bar{6}$$

При скелетизации необходимо удалять (стирать) отдельные элементы изображения объекта в окне сканирования. Условия стирания могут быть определены с учетом полученных признаков характерных точек окна:

- условие стирания верхнего элемента:

$$g_B = (\bar{1} \cup \bar{3} \cup \bar{0} \cup \bar{4}) \cdot \varphi_B \quad (1)$$

- условие стирания левого элемента:

$$g_L = (\bar{3} \cup \bar{5} \cup \bar{2} \cup \bar{6}) \cdot \varphi_L \quad (2)$$

- условие стирания правого элемента:

$$g_R = (\bar{7} \cup \bar{1} \cup \bar{6} \cup \bar{2}) \cdot \varphi_R \quad (3)$$

- условие стирания нижнего элемента:

$$g_H = (\bar{5} \cup \bar{7} \cup \bar{4} \cup \bar{0}) \cdot \varphi_H \quad (4)$$

Утоньшение состоит в последовательном преобразовании исходного символа на растре в новый символ путем стирания сначала крайних сверху, затем крайних слева, далее крайних снизу, а затем и крайних справа элементов.

Процесс продолжается до тех пор, пока все функции (1) - (4) не обратятся в единицу. Если одна из функций обратилась в единицу, то по этому направлению утоньшение прекращается.

Обычно для утоньшения символа требуется не более $b_{\max} / 2$ изменений, где b_{\max} - максимальная толщина линии в исходном символе.

1.6. Выделение характерных точек объекта

К характерным точкам объекта относится, в частности, центр тяжести (центр масс объекта). Определение его координат оказывается важным при отслеживании траектории перемещения объекта [1,7,15].

1. Определение координат центра масс объекта

$$\begin{cases} x_u = \sum_{i=1}^N x_i / N; \\ y_u = \sum_{i=1}^N y_i / N; \end{cases} \quad \text{для плоской фигуры}$$

С учетом "веса" отсчетов для многоуровневых ("трехмерных") изображений

$$x_u = \frac{\sum_j m_{ij} x_j}{M}; y_u = \frac{\sum_i m_{ij} y_i}{M}$$

$$M = \sum_i \sum_j m_{ij}$$

2. Заполнение контура (алгоритм штриховки или заливки)

- а) по критерию четности;
- б) по критерию "связности";
- 3. Определение координат центра тяжести объемного объекта:

$$x_c = \frac{\sum m_i x_j}{M}; y_c = \frac{\sum m_i y_i}{M}; z_c = \frac{\sum m_i z_i}{M}$$

$\sum m_i(x, y, z) \rightarrow$ статические моменты

Определение центра тяжести для однородных линий

$$\left. \begin{aligned} x_c &= \frac{\sum \Delta L_i x_i}{L}; \\ y_c &= \frac{\sum \Delta L_i y_i}{L}; \end{aligned} \right\} \text{ для плоских линий}$$

$$z_c = \frac{\sum \Delta L_i z_i}{L}; \quad L = \sum \Delta L_i - \text{длина линии}$$

Таким образом, в процессе обработки изображений осуществляется постепенное сокращение информационной избыточности и последовательный переход от естественного описания изображения к некоторому содержательному признаковому описанию, удобному для последующего распознавания объектов на изображении. Схематически такой процесс представлен на рис. 1.20.

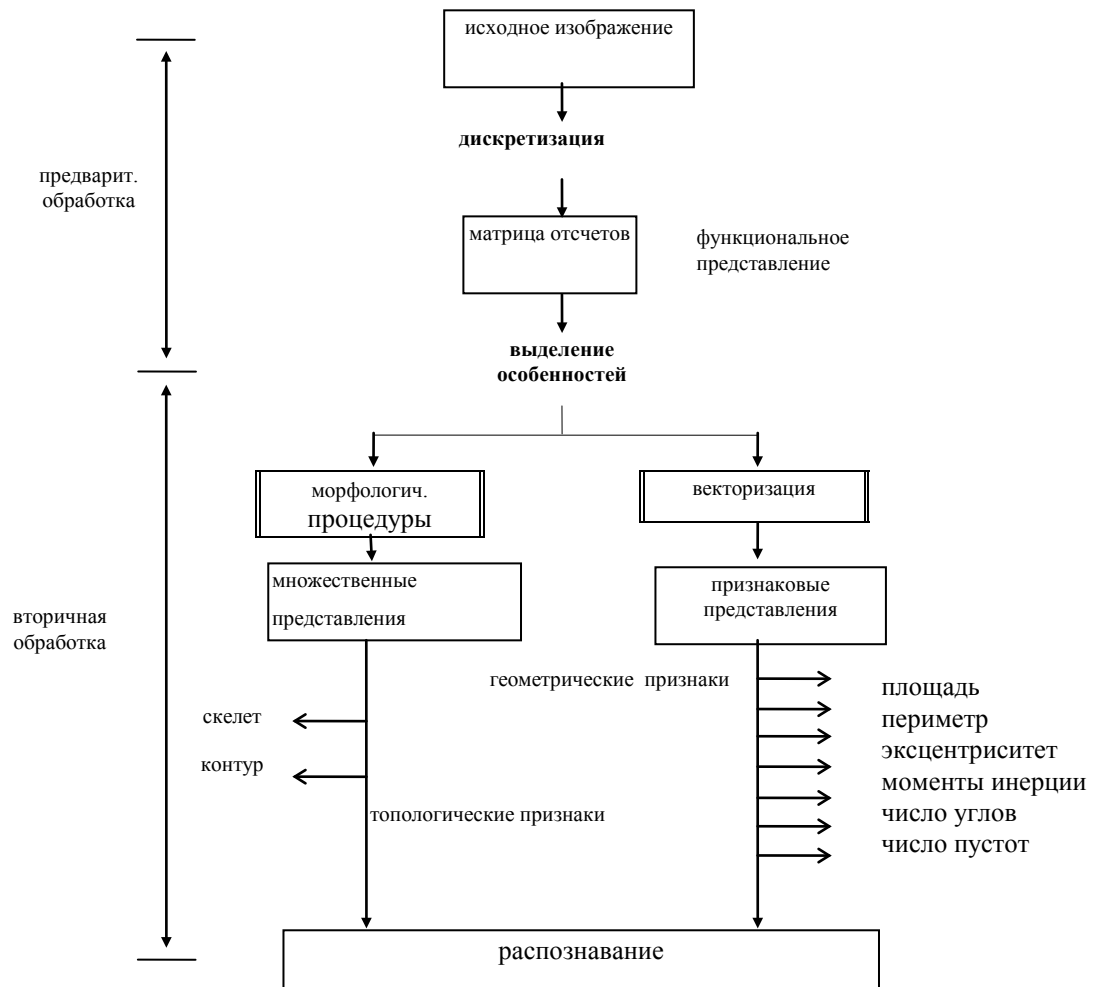


Рис.1.20. Переход от изображения к набору признаков при распознавании

2. ОСНОВЫ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ

За последнее десятилетие благодаря существенному росту информационной емкости запоминающих устройств и пропускной способности каналов связи в общем объеме информации, хранимой и передаваемой в информационно-телекоммуникационных системах, заметно возросла доля информации, приходящейся на изображения. При этом крайне актуальной стала проблема доступа к изображениям по их содержанию, для решения которой требуется распознавание изображений по эталону (в качестве которого может выступать как другое изображение, так и набросок от руки или словесное описание). Такая процедура требуется для организации поиска в специализированных базах изображений самого разнообразного назначения (базы логотипов, гео-информационные системы, базы биометрических данных, каталоги продукции, фотостоки и т.д.), а также для поиска произвольных изображений в домашних цифровых фотоальбомах, в сети Интернет и др.

Существует два основных подхода к проблеме извлечения изображений из баз данных [1]: с использованием текстовых аннотаций и путем распознавания изображений. Использование текстовых аннотаций является сравнительно простым в алгоритмическом плане и достаточно надежным средством, однако во многих случаях наличие текстовых аннотаций не может быть гарантировано, например, в силу большого числа изображений, их плохого вербализуемого содержания или по каким-либо другим причинам, и, кроме этого, при составлении таких аннотаций невозможно предсказать, на какие именно характеристики изображения будет направлен тот или иной запрос, поскольку аннотации оказываются весьма субъективными [2].

Проблеме распознавания изображений, начиная с середины прошлого века, было посвящено значительное число работ, однако до сих пор эта проблема далека от полного решения. Это связано с большой информационной емкостью и априорной неопределенностью, присущей изображениям (которые могут содержать произвольные объекты и быть сформированными с помощью сенсоров различных типов), а также с большой изменчивостью изображений (изменение ракурса или освещения приводит к изменению значений одновременно во всех элементах изображения) [3]. Кроме того, объекты, которые могут присутствовать на изображениях, обычно входят в сложные иерархические отношения типа "часть-целое" и "частное-общее".

Как результат, нет прямого соответствия между изображениями, представленными в информационных системах в виде массивов пикселей, и теми семантическими метками, которые

должны быть присвоены объектам на изображениях в процессе их распознавания [4]. Именно сложность решения общей задачи распознавания изображений ведет к разработке большого числа частных методов.

Из-за большого разнообразия существующих методов возникает вопрос об их сходстве и различии и эффективности применения при решении той или иной задачи. Одним из центральных компонентов методов распознавания принято считать используемое представление изображений [5]. Поскольку не удается построить явного отображения из пространства значений пикселей, составляющих изображение, в пространство семантических описаний сцены, оказывается необходимым вводить некоторые промежуточные уровни представления изображений. Под представлением обычно понимается формальная система, содержащая алгоритмы для получения в явном виде описаний объектов заданного класса [6, с. 36]. Тогда процесс распознавания объектов по их изображениям может интерпретироваться как процесс последовательного преобразования информации, начиная с исходного представления изображения в виде массива пикселей и заканчивая семантическим описанием изображения через некоторую совокупность промежуточных представлений.

Можно выделить следующие типы промежуточных представлений изображений [4]:

- низкоуровневые [7];
- признаковые [8, 9];
- контурные [10];
- структурные [3, 11].

В качестве конечных представлений в задачах распознавания изображений можно рассматривать представления, основанные на знаниях. В отдельную группу выделяются методы, основанные на знаниях [12]. Отличительной особенностью этих методов является то, что в них представление, основанное на знаниях, не только является выходным представлением системы распознавания изображений, но и используется в самом процессе распознавания. Иными словами, эти методы работают под управлением знаний с потоком информации «сверху вниз» в отличие от прочих методов, работающих под управлением данных с потоком информации «снизу вверх» [4].

В отдельную группу методов можно выделить иерархические методы [13], включающие одновременное использование нескольких промежуточных представлений перечисленных выше типов.

Помимо вопроса об используемых промежуточных представлениях изображений в процессе их распознавания возникает вопрос о том, каким образом строятся описания изображений в рамках выбранных представлений [5]. Здесь можно выделить

критерий качества описания, в наибольшей степени характеризующей методы распознавания, поскольку именно критерий качества определяет, какое именно описание изображения будет сформировано в рамках данного представления. Следующие подходы к заданию критерия качества являются наиболее типичными:

- эвристический (включающий самые разнообразные способы выбора решения, см., напр., [14, 15]);
- байесовский [16];
- энтропийный [17];
- теоретико-информационный [18].

Тип промежуточного представления изображений и подход к заданию критерия качества не являются исчерпывающими характеристиками метода распознавания изображений. К примеру, важен также и алгоритм оптимизации критерия качества [19], однако систематизация по данному компоненту на настоящий момент затруднительна ввиду отсутствия достаточной теоретической проработки проблемы.

Рассмотрим различные типы представлений и подходы к заданию критерия качества и проведем анализ возможности их применения в задачах распознавания изображений.

Однако, несмотря на многообразие, все методы распознавания можно разделить на три группы. Первая основана на непосредственном сравнении с эталоном, вторая — на понятии признаков и обработки в этом пространстве, а третья — на исследовании “конструкции” рассматриваемых образов (так называемое синтаксическое распознавание). Часто два первых класса методов объединяют в один.

2.1. Задача распознавания изображений

Выделим вопросы, которые возникают при решении задачи распознавания изображений:

- восприятие поля зрения,
- сегментация,
- нормализация выделенных объектов,
- распознавание.

Исходя из этого, используются следующие основные принципы:

- Принцип целостности - распознаваемый объект рассматривается как единое целое, состоящее из структурных частей, связанных между собой пространственными отношениями.

- Принцип двунаправленности - создание модели ведется от изображения к модели и от модели к изображению.

- Принцип предвидения. Заключается в формировании гипотезы о содержании изображения.

- Принцип целенаправленности, включающий сегментацию изображения и совместную интерпретацию его частей.

- Ничего не делать без процедуры понимания (восприятия поля зрения).

- Принцип максимального использования модели проблемной среды, использование заранее известных, априорных параметров.

Этап интерпретации изображения не обозначен четкими границами и включается частично в процесс сегментации и окончательно завершается на этапе распознавания.

Естественно задаться вопросом: а нельзя ли брать изображение и последовательно сравнивать его с эталонами по ряду каких-либо признаков? Но тут возникает ряд проблем и сложностей:

- **Фон.** Как правило, изображения предъявляются на сложном динамическом фоне.

- **Ориентация.** Изображения эталона и входные изображения отличаются положением в поле зрения.

- **Помехи.** Входные изображения не совпадают с эталонами за счет случайных и локальных помех.

- **Освещение.** Отличия входных и эталонных изображений возникает за счет изменения освещенности, подсветки.

- **Преобразования.** Эталоны и изображения могут отличать сложные геометрические преобразования

Для решения задачи в целом и на отдельных ее этапах применяются различные методы сегментации, нормализации и распознавания. На приведенной схеме указаны основные процедуры и методы обработки - от начального этапа восприятия поля зрения посредством датчиков до конечного, которым является распознавание. Кратко прокомментируем приведенную схему.

Предобработка

Процедура предварительной обработки используется практически всегда после получения информации с датчика, и представляет собой применение операций усреднения и выравнивания гистограмм, различного типа фильтров для исключения помех, возникающих в результате аппаратной дискретизации и квантования, а также подавления внешних шумов.

Сегментация

Под сегментацией будем понимать процесс поиска однородных областей на изображении. Наиболее часто применяются методы, основанные на определении однородных цветов или текстур,

однако для произвольной задачи этот этап не имеет четкого алгоритма.

При существовании стабильных различий в яркостях отдельных областей поля зрения применяются пороговые методы. Приведем пример: для сегментации методом порогового распределения необходимо получить бинарное изображение из полутонового. Для этого устанавливается некоторое пороговое значение. После квантования функция изображения отображает элементы изображения с уровнем яркости больше порогового в значение 1, меньшее порогового - 0.

При наличии устойчивой связности внутри отдельных сегментов эффективны методы наращивания областей. Этот принцип состоит в том, что происходит группировка соседних элементов с одинаковыми или близкими уровнями яркости, а затем объединение их в однородные области. Один из типов - центроидное связывание - предполагает выбор стартовых точек либо с помощью оператора, либо автоматически. Эффективным представляется метод водоразделов, основанный на поиске локальных минимумов с последующей группировкой вокруг них областей по связности.

Метод выделения границ хорошо применять, если границы достаточно четкие и стабильные. Выделение контурных линий наиболее часто используется в системах технического зрения и основано на учете изменения яркости и дальнейшем ее сравнении с пороговой.

Перечисленные методы служат для выделения сегментов по критерию однородных яркостей. Все перечисленные принципы приемлемы с точки зрения вычислительных затрат, тем не менее, для каждого из них характерна не единственность разметки точек в реальных ситуациях из-за необходимости применения эвристик.

Для описания и сегментации свойств изображений, к которым относятся однородность, шероховатость, регулярность, применяют текстурные методы, подразделяющиеся условно на две категории - статистические и структурные. Использование матриц совпадений, формируемых из исходных изображений, с последующим подсчетом статистических моментов и энтропии - есть основа статистического метода. При структурном подходе строится множество многоугольников и производится исследование на предмет общих свойств. Многоугольники с общими свойствами объединяют в области. На рис. 2.1 представлена последовательность основных процедур распознавания изображений и используемых при этом методов.



Рис.2.1. Основные процедуры и методы распознавания изображений

Распознавание

Перейдем к конечному этапу обработки изображения - распознаванию. Для этого этапа входными данными являются изображения, полученные в результате шумоподавления и процесса сегментации. Как правило, они отличаются от эталонных геометрическими и яркостными искажениями, а также сохранившимися шумами.

Для решения задач распознавания применяются, в основном, четыре подхода:

Корреляционный. Подход, основанный на принятии решений по критерию близости с эталонами. В основном применяется при обнаружении и распознавании изображений в системах навигации, слежения, промышленной роботизации. Наиболее трудоемкий подход с точки зрения потребления вычислительных ресурсов. Подразумевает под собой многошаговую корреляцию при полностью заданном эталоне, путем сканирования входного поля зрения. Другими словами, происходит перебор всех входных сигналов и сравнение их с эталонным.

Признаковый. Такие методы основаны на переходе в пространство признаков, а соответственно, требуют значительно меньших вычислительных мощностей. В зависимости от поставленной задачи, выполняется корреляционная обработка признаков, полученных от эталона и входного изображения. При этом возникает задача объединения и комплексной обработки признаков различной размерности (метрических, статистических, логических, текстурных и т. д.), полученных различными измерительными средствами с целью решения задачи распознавания.

Корреляционно-признаковый метод подразумевает под собой обработку статистическими методами признаков, полученных следующим образом. Изначально применяется метод частных корреляций для различных фрагментов эталонного изображения, а

затем в сигнальном пространстве полученные корреляционные коэффициенты рассматриваются как признаки.

Основной проблемой в признаковых методах составляет выбор признаков. При этом исходят из естественных правил:

- Признаки изображений одного класса могут различаться лишь незначительно (за счет влияния помех, шумов).
- Признаки изображений разных классов должны существенно различаться.
- Набор признаков должен быть минимальным (от их количества зависит надежность, сложность, скорость обработки).

Синтаксический метод основан на получении структурно-грамматических признаков, когда в изображении выделяются непроеизводные элементы - признаки. Вводятся правила соединения этих элементов, одинаковые для эталона и входного изображения. Анализ полученной таким образом грамматики обеспечивает принятие решений.

Каждый из подходов в распознавании имеет право на существование. Более того, в рамках каждого подхода есть свои конкретные алгоритмы, имеющие определенную область применения, которая зависит от характера различий входных и эталонных изображений, от помеховой обстановки в поле зрения, требований к объемам вычислений и скорости принятия решений. Признаковые и синтаксические методы - наиболее распространенные в теории распознавания образов.

Нормализация. Задача нормализации изображения - это задача определения параметров геометрических преобразований, которым подверглось изображение, с целью их компенсации. Компенсация может проводиться за счет изменения пространственного положения системы ввода изображения, либо алгоритмически путем применения обратного преобразования к входному изображению. Процедура преобразований производится с помощью операторов нормализации - нормализаторов, а вычисление параметров выполняется функциями, действующими на множестве изображений.

Методы нормализации при распознавании занимают промежуточное место между корреляционными и признаковыми алгоритмами. В отличие от признаковых, при нормализации изображение не исключается из рассмотрения, а только замещается изображением того же класса эквивалентности. В то же время, в отличие от корреляционных методов, множество входных изображений заменяется множеством нормализованных изображений. Каждая нормализованная картинка, вообще говоря, находится гораздо ближе к своему эталону (с позиции групповых преобразований), что значительно сокращает количество корреляций

на завершающем этапе распознавания.

Наибольший интерес в настоящее время в теории нормализации представляют последовательные методы, основанные на поэтапном вычислении параметров сложных преобразований и применении частичных нормализаторов на каждом этапе. На рис. 2.2. представлена общая схема процедуры распознавания изображений.



Рис. 2.2. Общая схема процедуры распознавания изображений

Дополнительные проблемы при решении задачи зрительного восприятия роботизированных систем по сравнению с традиционными задачами обработки и распознавания изображений:

- Описание среды функционирования. Необходимо комплексное описание на основе учета значительного объема априорной информации, создание модели проблемной среды, в отличие от задачи выделения конкретных признаков или выделения отдельных характеристик. Анализ 3D-объектов и учет законов перспективы. Необходимо учитывать не только проекции реальных объектов, но и проводить анализ в плане определения объемных пространственных отношений.
- Анализ множества произвольно расположенных объектов, выделение конкретных предметов при отсутствии или невозможности определить некоторые признаки (например, когда на плоскопараллельной проекции видна только часть контура

необходимого объекта, но уникальная и достаточная для его идентификации).

- Необходимость работы в реальной динамической среде. В общем случае, отсутствие постоянной задачи и необходимость оперативно реагировать на возникающие задачи.
- Необходимость согласованности при взаимодействии в реальном времени нескольких подсистем робота.

2.2. Методы распознавания на основе сравнения с эталоном

Представления изображений в виде необработанных данных

В задачах распознавания изображений в качестве исходного представления, из которого затем осуществляется отображение в конечное представление, обычно рассматривается представление изображения в виде массива «сырых» данных – набора результатов физических измерений, выполненных для некоторой сцены. Под сценой обычно понимается набор физических объектов или некоторый фрагмент реального мира [20, с. 11]. Чаще всего регистрируемую физическую величину представляет интенсивность электромагнитного излучения в некотором спектральном диапазоне, например, видимом, ИК, рентгеновском или радиодиапазоне. Однако изображение может содержать и совершенно другую информацию и представлять собой, например, цифровую карту высот, гидроакустические данные или распределение плотности некоторых элементарных частиц в зависимости от направления.

Элементом изображения при начальном представлении является пиксель (от pixel: picture element), поэтому такие представления обычно называют представлениями на уровне пикселей [21].

Пиксели, образующие изображение, с соответствующими им значениями интенсивностей организованы в массив, размерность которого определяется исходной природой данных. Упорядочивание пикселей в массиве соответствует пространственной организации сцены. Для многих задач массив пикселей является двумерным, но широко распространены приложения, например, медицинские, в которых размерность массива может быть больше двух. В случае объемных изображений их элементы принято называть вокселями (от voxel: volume pixel). Последовательность изображений может трактоваться как изображение большей размерности (на уровне пиксельных представлений время ничем не отличается от дополнительной пространственной координаты).

Основной проблемой, связанной с представлениями на пиксельном уровне, является проблема эффективности этих представлений в задачах хранения и передачи информации, что особенно актуально для объемных изображений. Существует несколько альтернативных способов представления объемных изображений [4], такие как, например, трехмерный массив вокселей, стереопара, проекция на Гауссову сферу и т.д. Однако разработка таких представлений вызвана необходимостью решения чисто технических проблем, не связанных с задачами распознавания изображений, поэтому здесь мы на них останавливаться не будем.

Представления на пиксельном уровне являются исходными для любых приложений интерпретации изображений в информационно-телекоммуникационных системах. Именно потому, что это представление является общим для различных задач компьютерной обработки изображений, часто говорят, что "изображения – это массивы пикселей" [4]. Однако стоит отделять изображение (содержание) от его конкретного представления (формы).

Представления на пиксельном уровне содержат в себе всю имеющуюся информацию о наблюдаемой сцене, но в форме, неудобной для автоматического анализа [20, с. 12]. В частности, такие представления редко используются для распознавания или сопоставления изображений. Это и вызывает необходимость привлечения других представлений изображений с целью извлечения содержащейся в них релевантной информации.

Низкоуровневые представления

Представления на пиксельном уровне не содержат информации о том, как следует работать с изображениями. Для определения набора операций над изображениями естественно их представить в качестве элементов некоторого математического пространства. Это дает обширный набор строгих внутренне непротиворечивых средств анализа и преобразования изображений. В частности, если выбранное пространство является метрическим, то появляется возможность формального определения "расстояния" (т.е. степени сходства) между изображениями, что является основой для их распознавания.

Два основных класса математических моделей изображений включают [7]: функциональные представления и представления в виде случайных полей.

Функциональные модели

При использовании функциональных моделей изображение интерпретируется как функция из некоторого (например, Гильбертова) пространства:

$$f: G \rightarrow V, G \subseteq R^n, V \subseteq R^m,$$

где G – область определения функции, а V – область ее значений. Величина n – размерность изображения (обычно равная двум), а m – размерность вектора значений, хранящихся в каждом пикселе. Например, для полутоновых изображений $m=1$, а для цветных RGB изображений $m=3$.

Функциональное представление является базовым для проведения таких операций над изображениями, как пространственное преобразование изображений, преобразования интенсивности, фильтрация и др.

Пространственное преобразование изображения (масштабирование, вращение и т.д.) осуществляется с помощью преобразования координат в области G :

$$f_2(\mathbf{x}') = f(g(\mathbf{x})), g: G \rightarrow G_2,$$

где g – функция, ставящая в соответствие каждой точке из области определения G исходного изображения f точку в области определения G_2 преобразованного изображения f_2 . Преобразование интенсивности (например, изменение яркости или контраста для всего изображения) или преобразования цвета (если V – это цветовое пространство) может быть представлено следующим образом:

$$f_2(\mathbf{x}) = h(f(\mathbf{x})), h: V \rightarrow V_2,$$

где h – функция, зависящая только от значения функции f в данной точке и ставящая в соответствие этому значению новое значение, возможно, из другой области V_2

Часто привлекаемой операцией над изображениями является линейная фильтрация, которая представима в виде свертки с некоторым ядром:

$$f_2 = T(f) = f \otimes \varphi, (f \otimes \varphi)(\mathbf{x}) = \int_G f(\mathbf{y})\varphi(\mathbf{x}-\mathbf{y})d\mathbf{y}.$$

К примеру, с помощью линейной фильтрации представляются такие операции, как сглаживание изображения или его дифференцирование.

В ряде работ [20, 21], посвященных проблеме автоматического распознавания изображений, отмечается полезность применения спектрального описания изображений.

Под двумерным пространственным спектром изображения $f(x,y)$ понимается комплексная в общем случае функция $F(u, v)$, связанная с $f(x,y)$ следующим соотношением:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp[-i(ux + vy)] dx dy \quad (2.1)$$

где N – число пикселей на изображении, u и v – пространственные частоты.

Такое описание изображений обладает существенными преимуществами, полезными при распознавании.

1. Модуль спектральной функции (2.1): $|F(u, v)|$ не зависит от переноса функции $f(x, y)$, т.е. описание является инвариантным к переносам изображения в плоскости наблюдения.

2. Описание изображений (2.1) обладает определенной помехоустойчивостью. В тех случаях, когда спектры распознаваемого изображения и аддитивной помехи различны, можно увеличить отношение сигнал/шум с помощью пространственной фильтрации.

3. Поворот изображения вокруг произвольной точки приводит к повороту пространственного спектра $F(u, v)$ вокруг начала координат с соответствующим изменением фаз составляющих (изменение фаз не влияет на $|F(u, v)|$). Поэтому при необходимости устранения ориентации распознаваемого изображения на спектральное описание, следует проинтегрировать модуль спектра, записанный в полярных координатах, по соответствующему углу в пределах от 0 до 2π . (Если такая операция осуществляется в оптической системе с предварительной регистрацией пространственного спектра на фотоносителе, то это соответствует переходу к описанию изображения его автокорреляционной функции). Следует, однако, учитывать, что такое интегрирование спектра может привести к значительным потерям информации от изображения.

4. Если изображение $f(x, y)$ имеет спектр $F(u, v)$, то изображение $f(ax, ay)$, связанное преобразованием подобия с $f(x, y)$, где a - постоянный коэффициент, имеет спектр

$$\frac{1}{a^2} F\left(\frac{u}{a}, \frac{v}{a}\right)$$

Несмотря указанные преимущества, вытекающие из использования пространственных спектров изображений, трудности, связанные с вычислением спектров на ЭВМ, не позволили использовать спектральное описание для распознавания двумерных изображений.

Распознавание изображений в рамках функциональных моделей

Задача распознавания изображений, интерпретируемых как функции, сводится к поиску пространственного преобразования и, возможно, преобразования яркости, которые минимизируют расстояние между изображениями в заданном метрическом пространстве. Следует также отметить, что модели задаются обычно в одной из следующих форм:

1. в виде жестких (“неподвижных”) эталонов - детерминистская модель;
2. в виде гибких (“подвижных”) эталонов, задаваемых с точностью до неопределенных параметров или группы непрерывных преобразований - квазидетерминистская модель;
3. в виде статистических распределений множества параметров первичного описания (скажем, дискретных отсчетов, коэффициентов Фурье и т.п.) - статистическая модель.

Следует отметить, что процесс обучения может состоять в получении оценок для упомянутых выше неопределенных параметров или распределений. Таким образом, резкой границы между двумя рассматриваемыми методами распознавания не существует.

Выбор той или иной модели определяется особенностями классов распознаваемых изображений, а также соображениями достоверности правильной классификации и простоты реализации. При наличии достаточно полной модели классов решающее правило может быть относительно легко найдено. Так, например, в случае детерминистской и квазидетерминистской моделей естественно сначала определить близость опознаваемого изображения (f) до эталонных (f_{i_0}). Для этого формируются некоторые функционалы - меры близости $\rho(f ; f_{i_0})$. Наиболее часто используются следующие меры:

$$\rho_i^{(1)} = \max |f(x, y) - f_{i_0}(x, y)| \quad (2.1)$$

$$\rho_i^{(2)} = \left\{ \frac{1}{S_D} \iint_{(D)} [f(x, y) - f_{i_0}(x, y)]^2 dx dy \right\}^{\frac{1}{2}} \quad (2.2)$$

$$\rho_i^{(3)} = \frac{\left[\iint_{(D)} f(x, y) f_{i_0}(x, y) dx dy \right]^2}{\iint_{(D)} f(x, y) dx dy \iint_{(D)} f_{i_0}^2(x, y) dx dy} \quad (2.3)$$

Первые две меры характеризуют отклонение от эталона, а третья является коэффициентом взаимной корреляции. Если используются подвижные эталоны, зависящие от параметров $\{a_{i_0}\}$, то в качестве мер близости естественно задавать минимальные значения $\rho_i^{(1)}$ и $\rho_i^{(2)}$, но максимальное значение $\rho_i^{(3)}$, достигается путем вариации этих параметров. Естественно, что меры (2.1)-(2.3) могут определяться не только в функциональном пространстве F первичного описания, где изображения задаются функциями f распределения яркости, амплитуды или фазы света, но и в векторном

пространстве X первичного описания, когда изображения описываются семейством параметров $\{x_k\}$ - “первичных признаков”.

Использование этой нормы ведет к широко распространенным корреляционным методам [23, 25, 61], в которых степень сходства двух изображений определяется как корреляция. При разных взаимных сдвигах изображений взаимную корреляционную функцию удобно вычислять через обратное дискретное преобразование Фурье (ДПФ) произведения Фурье-спектров:

$$C_{f_1, f_2}(x, y) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \tilde{f}_1(u, v) \tilde{f}_2^*(u, v) \exp \left[i \frac{2\pi}{N} (ux + vy) \right]. \quad (2.4)$$

Простейшие корреляционные методы могут быть использованы для поиска смещения между эталонным и распознаваемым изображениями при условии, что интенсивности изображений преобразованы линейно. Несмотря на существенные ограничения, накладываемые на возможные преобразования изображений, классический корреляционный метод широко используется для решения ряда частных задач.

Рассмотрим подробнее метод оконной обработки спектра. Это метод основан на систематическом изучении свойств пространственных спектров и их связи с исходным фотографическим изображением, т.е. тех свойств пространственных спектров, которые соответствуют важным с точки зрения конкретной задачи распознавания свойствам исходных фотографических изображений.

Изучив такие свойства пространственных спектров, можно выбрать способ (геометрию) окон, сохраняющий необходимую нам информацию и уменьшающий объем информации, подлежащей передаче в решающее устройство. Выбор окон (фрагментов) пространственного спектра сводится к измерению количества света в пределах заданных участков пространственного спектра. Это можно осуществить с помощью набора окон, покрывающих в совокупности весь пространственный спектр. Каждое окно помещается в частотную плоскость и свет, проходящий через окно, измеряется и преобразуется в число. Набор таких чисел поступает в решающее устройство. Из литературы известен ряд наборов спектральных окон различной конфигурации пригодный в основном для выявления регулярностей (периодичность одномерная и двумерная, линейная структура) в изображении [4, 23].

Интегрируя световую энергию вдоль окружности с центром на оптической оси, мы получаем полный вклад одной частотной составляющей спектра независимо от направления.

В действительности, окна в виде концентрических колец имеют конечную ширину и учитывают вклад не одной, а нескольких

частотных составляющих. Причем конечная ширина окна выгодна тем, что уменьшает чувствительность выборки к незначительным изменениям в изображении (за счет шумов). Набор таких окон (рис. 2..3, а-в) позволяет получить набор величин,

$$m_j = \int_0^{2\pi\theta_j + \Delta\theta} \int_{\rho_j} F(\rho, \theta) \rho d\rho d\theta,$$

$$j=1,2,\dots, \rho = \sqrt{u^2 + v^2}, \theta = \arctg \frac{u}{v}$$

по которым можно судить о наличии одно- или двумерной периодичности в изображении.

Радиальная линия в частотной плоскости соответствует единственному направлению в изображении, которое включает все частотные компоненты. Набор величин

$$m_j = \int_{\rho_{\min}}^{\rho_{\max}} \int_{\theta_j}^{\theta_j + \Delta\theta} F(\rho, \theta) \rho d\rho d\theta$$

характеризует все направления, если $m_j(j=1,2,\dots,n)$ покрывают сектор от 0 до 360^0 . На практике используется окно фрагментации клинообразной формы. Такое окно определяет вклад от небольшого числа смежных направлений и обладает тем преимуществом, что уменьшает число необходимых выборок и снижает влияние небольших изменений. Набор фрагментирующих окон клинообразной формы позволяет получить описание амплитудного спектра вдоль радиального направления. Этот метод дискретизации нечувствителен к масштабу изображения.

Заметим, что по существу фрагментации спектра представляет собой частный случай метода обобщенных моментов амплитудного спектра.

Таким образом, представляется целесообразным применение спектрального описания изображений при их распознавании по методу обобщенных моментов. При этом важнейшим преимуществом спектрального описания является его инвариантность к сдвигу.

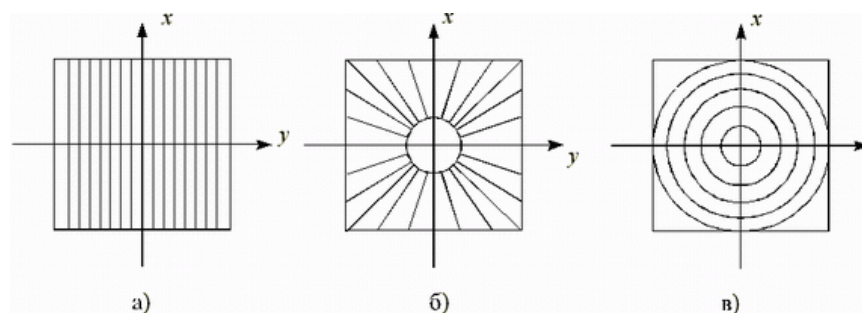


Рис. 2.3. Набор оптических окон, используемых при обработке пространственно-частотных спектров

Существуют более сложные модификации корреляционного метода, к которым можно отнести локальную корреляцию [13, 25] и метод Фурье-Меллина [45, 64].

В методе локальной корреляции производится поиск смещений каждой точки эталонного изображения относительно распознаваемого с использованием корреляции по некоторой окрестности точки. Это позволяет восстанавливать нелинейные пространственные преобразования двух изображений, что может использоваться, например, в задачах стереозрения или распознавания объектов с изменяющейся формой. Однако метод локальной корреляции имеет свои ограничения, например, требование к отсутствию глобального преобразования координат (в том числе вращения и масштабирования) и интенсивностей.

В методе Фурье-Меллина используется инвариантное к сдвигу преобразование изображения (модуль Фурье-спектра). Затем осуществляется переход в полярно-логарифмическую систему координат, в которой масштабирование и поворот исходного изображения эквивалентно сдвигу вдоль соответствующих осей. Метод Фурье-Меллина позволяет распознавать изображения, полученные из эталона путем применения к ним преобразования из группы подобия.

Функциональные представления изображений, как правило, инвариантны к сравнительно простым преобразованиям изображений (зачастую к пространственным преобразованиям ограниченных классов), поэтому соответственно ограничена сфера их применения.

Один из основных способов обнаружения объектов на изображении состоит в сравнении с эталоном. При этом эталон интересующего нас объекта сравнивается со всеми неизвестными объектами, находящимися на изображении [45].

Процедура сравнения состоит в вычислении функции взаимокорреляции объекта с эталоном, согласно выражению:

$$y_{i,j}^{(k)} = \sum_{n=1}^M \sum_{m=1}^M g_{mn}^{(k)} * x_{i+m-1, j+n-1},$$

$$Y = \max_{ij} \{ y_{ij}^{(k)} \}$$

где $g_{mn}^{(k)}$ - отсчеты k-го эталона, причем значение служит критерием соответствия исходного объекта г-му эталону. В случае если отсчет и эталоны являются бинарными, процедура сравнения выполняется по битовым картам.

Пусть изображение исходного символа задано в виде двумерного бинарного массива B_N ,

$b_{mn} = \{0,1\}$, $m = 1, \dots, N$, $n = 1, \dots, N$, а набор эталонов задан в виде матриц $W(k)$, $k = 1, \dots, K$, (K - число эталонов). Тогда степень сходства исходного символа с одним из эталонов можно определить так:

$$G_{ij}^{(k)} = \sum_n^M \sum_m^M b_{i+m-1, j+n-1} \otimes w_{mn}^{(k)}$$

Здесь символ \otimes соответствует операции неравнозначности (отрицания сложения по mod2).

Если исходный символ соответствует эталону, то функция сходства $G_{ij}^{(k)}$ достигает максимума. Может быть заранее задан порог чувствительности, определяющий минимальное значение функции сходства, при котором принимается решение о соответствии исходного символа некоторому эталону. В результате сравнения со всеми эталонами будет определен один или несколько эталонов, для которых превышен порог чувствительности. Поэтому может потребоваться дополнительный этап, на котором будет принято окончательное решение об идентификации исходного символа. Это может быть произведено путем определения абсолютного максимума функции сходства.

2.2.1. Распознавание по проекциям

Для сокращения объема вычислений при определении функции сходства корреляционно-разностный метод может быть модифицирован. В частности, определение максимума функции сходства может выполняться по проекциям битовых карт символа и эталона.

Под проекцией символа (или эталона) на ось X понимается вектор, каждый элемент которого равен сумме единичных бит в соответствующем столбце битовой карты, а под проекцией на ось Y - вектор, каждый элемент которого равен сумме единичных бит в соответствующей строке битовой карты:

$$\text{Pr } X = [prX_0, prX_1, prX_2, \dots, prX_M]$$

$$\text{Pr } Y = [prY_0, prY_1, prY_2, \dots, prY_M]$$

$$\text{Pr } X_i = \sum_j^1 b_{i,j}; \quad \text{Pr } Y_j = \sum_i^1 b_{i,j}$$

Функция сходства проекций определяется как:

$$F^X(k) = \sum_i | \text{Pr } X_i - \text{Pr } X_i^k |$$

$$F^Y(k) = \sum_j | \text{Pr } Y_j - \text{Pr } Y_j^k |$$

где $\text{Pr } X^k, \text{Pr } Y^k$ - проекции k -го символа-эталона.

В результате сравнения проекций исходного символа с проекциями всего набора эталонов выделяется один или несколько эталонов, имеющих проекции, близкие к проекциям исходного символа. Следующий этап распознавания может заключаться в сравнении битовых карт исходного символа с битовыми картами только тех эталонов, которые были выделены на первом этапе.

Подобный подход может быть использован и для распознавания символов, описываемых матрицей малоразрядных двоичных целых чисел. В этом случае осуществляется сравнение по разрядным битовым картам, что позволяет вычислить функции $F^q(k)$, где $q = 0, \dots, Q - 1$ определяет q -ю разрядную функцию, Q - число разрядов в описании символа и эталона. Результирующая функция сходства определяется из соотношения [45]:

$$F(k) = \sum_{q=1}^Q F^q(k) 2^q \quad (2.5)$$

Если сходство между неизвестным объектом и эталоном достаточно велико, то этот объект помечается как соответствующий эталонному объекту. Полное совпадение эталона с какой-либо частью изображения бывает редко из-за действия шумов и искажений, а также вследствие отсутствия первоначальной информации относительно точной формы и структуры объекта, который требуется обнаружить. Поэтому указывают на наличие выделенного объекта там, где различие меньше всего.

Метод обладает высокой точностью распознавания для заданного набора эталонов даже при наличии случайного шума. Важным достоинством метода является то, что эталоны задаются непосредственно в виде битовых карт символов, поэтому не требуется дополнительных затрат на подготовку эталонов.

Главный недостаток метода сопоставления с эталоном заключается в необходимости использования огромного количества эталонов для учета изменений объектов, возникающих при их повороте и увеличении (уменьшении) размеров.

2.2.2. Распознавание на основе анализа двумерных гистограмм

Гистограммы представляют собой распределение пикселей изображения по яркости (по уровням), т.е. описывают частоту появления значений отдельных элементов (пикселей) независимо от других.

Очевидно, что значения частоты не зависят от пространственного распределения элементов изображения. Иначе говоря, вид гистограммы не изменится при повороте или сдвиге объекта.

Дальнейшим развитием гистограмм являются гистограммы второго порядка, которые характеризуют частоту совместного появления двух одновременно анализируемых объектов.

Для бинарных изображений возможны следующие двумерные гистограммы: типа 00, 01, 10 и 11.

$$\left. \begin{aligned} h_{mn}^{00} &= \sum_i \sum_j \bar{b}_{ij} \& \bar{b}_{i-m, j-n}; \\ h_{mn}^{01} &= \sum_i \sum_j \bar{b}_{ij} \& b_{i-m, j-n}; \\ h_{mn}^{10} &= \sum_i \sum_j b_{ij} \& \bar{b}_{i-m, j-n}; \\ h_{mn}^{11} &= \sum_i \sum_j b_{ij} \& b_{i-m, j-n}; \end{aligned} \right| i-m, j-n \in I \times J$$

$$H^{11} = [h_{mn}^{11}]$$

Такие гистограммы получили название гистограммы векторов расстояний и представляют собой двумерные матрицы, что затрудняет работу с ними.

Поэтому для удобства работы с двумерными гистограммами строят гистограммы модулей и направлений которые представляют из себя вектора.

Гистограмма модулей может быть определена как:

$$h_{\text{mod } m, n}^{11} = \sum_m \sum_n h_{mn}^{11} \left| \sqrt{m^2 + n^2} \right| \in (r_k, r_{k+1})$$

где $\{r_k, r_{k+1}\}$ - k -тый подинтервал.

В свою очередь, гистограмма направлений описывается следующим образом:

$$h_{dl}^{11} = \sum_m \sum_n h_{mn}^{11} \left| \arctg\left(\frac{m}{n}\right) \right| \in (\varphi_l, \varphi_{l+1})$$

Процесс формирования подобных гистограмм направлений и модулей иллюстрирует рис. 2.3.

Основные свойства двумерных гистограмм:

1. H^{11} симметричны относительно начала координат.
2. H^{11} , H_{mod}^{11} и H_d^{11} не зависят от сдвига изображения.
3. H_{mod}^{11} не зависит от угла поворота.
4. Вращение бинарной фигуры приводит к циклическому сдвигу отсчетов гистограммы H_d^{11} .

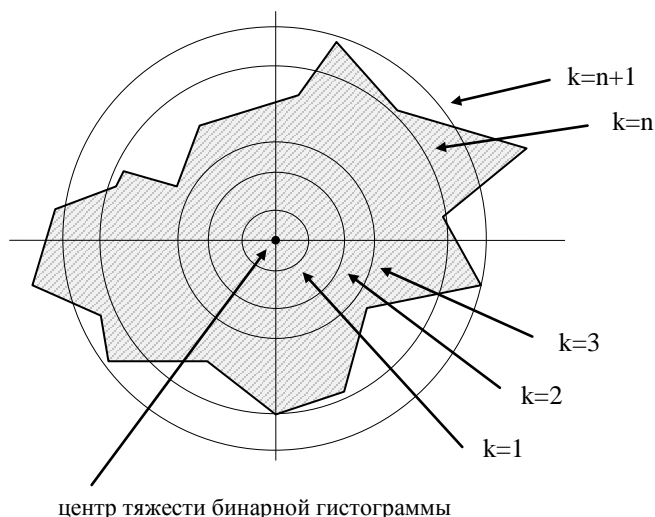


Рис. 2.3. Формирование гистограмм модулей и направлений

2.3. Признаковые методы распознавания

Признаковые представления

Признак определяется как функция от значений, содержащихся в одном или более пикселях, и вычисляется так, что численно выражает некоторую значимую характеристику объекта [45].

Можно классифицировать различные признаки изображений следующим образом [4]:

1. Общие признаки: признаки, независимые от приложения, такие как цвет, текстура, форма. Они могут быть далее разделены на:

- признаки пиксельного уровня: признаки, вычисляемые в каждом пикселе, такие как цвет, положение;
- локальные признаки: признаки, вычисляемые в некотором окне или ограниченной области изображения;
- глобальные признаки: признаки, вычисляемые по всему изображению. Обычно, это статистические свойства изображений, например, гистограмма, среднее значение, дисперсия и другие статистические моменты.

2. Предметно-зависимые признаки: признаки, зависящие от приложения, например, описывающие лица людей, отпечатки пальцев и т.д. Эти признаки формируются на основе общих признаков для конкретной предметной области.

С другой стороны, все признаки могут быть условно разделены на низкоуровневые и высокоуровневые признаки. Низкоуровневые признаки могут быть извлечены непосредственно из исходного изображения, тогда как высокоуровневые признаки базируются на низкоуровневых признаках.

Точки, в которых значение некоторого локального признака достигает экстремального значения, называются точками интереса [45]. Однако описание изображения в виде совокупности взаимосвязанных точек интереса следует рассматривать как описание следующего уровня абстракции, о чем, в частности, свидетельствует и то, что методы распознавания изображений на основе точек интереса во многом сходны со структурными методами.

Выбор в качестве первичных признаков распознаваемого изображения тех или иных физических величин определяется:

1. инвариантностью признаков к всевозможным преобразованиям изображения;
2. простотой устройства измерения выбранных признаков и устройства, преобразующего полученный параметр в код;
3. чувствительностью признаков к мешающим воздействиям, т.е. помехоустойчивостью;
4. сложностью обучения системы.

В каждой конкретной задаче важен правильный выбор “наиболее информативных признаков” распознаваемого изображения, так как он определяет лучшую конструкцию устройства и более широкие перспективы его использования.

В самом общем случае признаками распознаваемого изображения является набор чисел, которые согласно классическому представлению, являются координатами вектора объекта (изображения) в n -мерном пространстве признаков (n - число признаков). Переход от двумерной функции (изображения) к системе чисел (признаков) можно интерпретировать как задание некоторого функционала.

**Распознавание изображений на основе признаков **

Считая этот функционал линейным, можно аналитически записать процесс извлечения признаков в виде скалярного произведения

$$x_k = (f, \varphi_k) \quad (2.5)$$

где: φ_k - некоторая система функций; f - изображение; x_k - количественная мера признака (число). Такое описание процесса извлечения признаков было впервые дано А.А.Харкевичем [48].

Определяя x_k как

$$x_k = \iint f(x, y) \varphi_k(x, y) dx dy \quad (2.6)$$

очевидно, что выбор признаков сводится к выбору системы функций $\varphi_k(x, y)$, ($k = 1, 2, \dots, N$). Выражение (1.8) формально совпадает с определением обобщенных моментов [15, 16] функции f относительно

системы весовых функций φ_k . Частными случаями (2.6) являются: дискретные выборки (“рецепторное поле”), коэффициенты обобщенного ряда Фурье (случай ортогональных функций φ_k), степенные моменты ($\varphi_{jk} = x^j y^k$), выходы бинарных фильтров ($\varphi_k=1$, при $(x,y) \in G_k$ и $\varphi_k=0$ при $(x,y) \notin G_k$), коэффициенты разложения в обобщенный ряд Карунена-Лоэва [13] и т.п.

Почти все известные методы получения признаков [45] фактически являются частными случаями метода обобщенных моментов. Для задач распознавания случайных изображений в некоторых случаях может быть использован метод допустимых преобразований, позволяющий определить универсальное понятие сходства распознаваемого изображения и эталона. В этом случае предполагается, что изменения претерпевают эталонные изображения в определенной заданной области. Эта область эталонов, представляющая собой множество образов, получается из основного эталона с помощью определенных допустимых линейных преобразований. Этот метод учитывает две характерные трудности большинства задач по распознаванию – большое разнообразие образов и наличие случайных помех. Разнообразие распознаваемых образов учитывается путем описания множества эталонных образов в виде функции $E(k,b)$, линейно зависящей от мешающих параметров k и b . Помехоустойчивость метода обусловлена тем, что при распознавании не требуется точного совпадения распознаваемого изображения с одним из эталонов, так как алгоритм распознавания сводится к отысканию максимума некоторой функции, выражающей сходство изображения и эталона по параметрам, от которых зависит эталон. Однако даже при сравнительно небольшом числе параметров отыскание максимума сходства представляет большие вычислительные трудности и требует метода динамического программирования.

Следует заметить, что этот метод по своей сущности близок к методу, основанному на использовании большого числа эталонных изображений. Действительно, можно трактовать совокупность эталонных изображений класса как множество “дискретных значений” некоторого динамического эталона, с которым производится сравнение распознаваемого образа.

Если классы распознаваемых изображений имеют характерные фрагменты $\theta_k(x', y')$, то целесообразно их использование для формирования признаков:

$$m_k = \iint_S f_k(x, y) \varphi_k(x, y) dx dy \quad (2.7)$$

где $f_k(x, y) = \theta_k(x', y')$.

Обозначим области, занимаемые характерными фрагментами

изображений, через S_k и будем использовать бинарные функции $\theta_k(x', y')$ вида

$$\theta_k(x', y') = \begin{cases} 1, & \text{если } (x', y') \in S_k \\ 0, & \text{если } (x', y') \notin S_k \end{cases} \quad (2.8)$$

для формирования признаков:

$$m_k = \iint_{S_k(x', y')} f(x, y) dx dy \quad (2.9)$$

Заметим, что этот важный вид признаков весьма просто реализуется в оптике путем использования диафрагм с фигурными окнами (для реализации m_k достаточно спроектировать изображение на диафрагму с окном, соответствующим области S_k и проинтегрировать с помощью фотоприемника проходящий через окно световой поток).

Выбор области задания фрагментов (S_k) определяется эвристически, а способы реализации описаны в [13].

Общим соображением при выборе S_k является следующее: желательно выбирать такие подмножества фрагментов, которые с одной стороны хорошо описывают данный класс, а с другой стороны – редко встречаются в изображениях остальных классов.

Для получения признаков могут быть использованы коэффициенты разложения по полным ортогональным системам функций $\{\varphi_{ik}\}$ [13, 40]. К ним следует отнести тригонометрические многочлены, полиномы Лежандра, полиномы Чебышева, а также функции Радемахера, Хаара, Уолша и некоторые другие. Также могут быть использованы для формирования признаков квазиортогональные системы функций, например, двумерные бинарные случайные функции. Такие функции обладают с точки зрения формирования признаков одним существенным преимуществом по сравнению с функциями Уолша, Радемахера, Хаара и др., которые характеризуются увеличением частоты элементов структуры с ростом номера реализации функции (т.е. изображение становится все более и более мелкоструктурным). Поэтому, с одной стороны при использовании функций Уолша структура изображения оказывается не связанной со структурой оптической маски с записью реализации используемой системы функций, а с другой стороны при необходимости большого числа признаков реализации функций типа Уолша высоких порядков не обеспечат достоверное измерение признаков вида (2.7), поскольку различие между величинами соседних признаков будет соизмеримо с ошибками измерения в оптике.

В то же время использование случайных бинарных функций

позволяет выбрать частоту дискрета в соответствии с размером элементов изображения и избежать указанных трудностей.

Важным достоинством этих систем является универсальность получаемой системы признаков, основанная на полноте систем функций (при увеличении числа признаков точность описания изображений неограниченно возрастает). Однако за это качество приходится расплачиваться неэкономичностью описания - требуемое число признаков может быть весьма значительным (до нескольких десятков). В этом смысле признаки типа коэффициентов разложения по полным системам $\{\varphi_k\}$ противоположны “фрагментам” (2.9). Признаки типа “фрагментов” не образуют полной системы, не универсальны, но зато могут давать чрезвычайно экономное описание.

Описание некоторого изображения или каких-то его областей совокупностью признаков позволяет применять дискриминантные методы распознавания образов, поэтому признаковый подход наиболее широко используется в распознавании объектов [27, 83] или текстур [39, 40], принадлежащих ограниченному числу классов, то есть в условиях низкой априорной определенности.

Признаковые методы применяются в целях распознавания только в тех случаях, когда удастся построить инвариантные признаки. При этом в большинстве случаев построение таких признаков для конкретных условий является задачей разработчика

После того как признаки выбраны, для решения задач распознавания изображений может быть использован математический аппарат распознавания образов. В частности, популярными являются такие методы, как метод опорных векторов [3] и метод на основе смесей нормальных плотностей распределения вероятностей [3].

Использование инвариантных признаков позволяет осуществить распознавание объектов или категоризацию (например, разделение по сюжетам) изображений в достаточно широком круге приложений, в частности, в задачах распознавания [4, 100], однако описание сложных сцен или объектов, подверженных сложным типам изменчивости, в рамках признакового подхода оказывается затруднительным, как и распознавание по глобальным признакам объектов по их фрагментам.

В основе методов данной группы лежит гипотеза о представлении объекта (образа) в виде вектора признаков S , принадлежащего множеству V . Тогда множество образов представляют в виде множества векторов S , разбитого на N подмножеств, таких, что каждый вектор, отнесенный в результате классификации к j -му классу, принадлежит подмножеству E_j [45].

Иначе говоря:

$$\bigcup_{i=1}^M E_i \equiv V; \quad E_i \cap E_j = 0 \quad (\forall i, j \in \mathbb{N} \text{ и } i \neq j)$$

Последнее свойство означает свойство разделения множества на непересекающиеся подмножества. В этом случае, если обеспечивается такое свойство, множество V обладает сепарабельным разделением. Однако на практике подобное свойство обеспечивается крайне редко, что позволяет считать разбиение частично-сепарабельным.

Операция классификации состоит в том, чтобы распределить данные по категориям (или классам), где под классом понимают совокупность образов, имеющих одни и те же признаки.

Пусть X - множество объектов, $X = [x_1, x_2, x_3, \dots, x_N]$, причем каждый объект обладает множеством признаков $C = [c_1, c_2, c_3, \dots, c_p]$. Разбиение исходного множества на классы (подмножества) объектов в зависимости от совокупности значений признаков считается завершенным, если $X_i \cap X_j = 0$, если $\forall i, j = \overline{1, n}; i \neq j$ и $X = \bigcup_{i=1}^n X_i$, где $n \leq N$

- число классов объекта. Тогда говорят, что подмножества X_i ($X = \bigcup_{i=1}^n X_i$)

- есть результат разбиения $\Pi(x)$ исходного множества на классы.

Задача заключается в определении функции $f: X \rightarrow \Pi(x)$, обеспечивающей такое разбиение. Признаки или свойства объекта могут быть подразделены на следующие группы:

1. Количественные признаки, когда значение признака получено в результате измерения определенной физической величины (массы, температуры и т.д.);
2. Вероятностные признаки, которые характеризуются тем, что каждому свойству или наличию каждого признака сопоставляется вероятность появления этого признака у объекта;
3. Логические признаки (принципиальное наличие или отсутствие признаков).

На практике чаще всего используются наборы, сочетающие все указанные типы признаков, а для распознавания используются только существенные признаки $|Y| \in |C|; |Y| < |C|$.

Процедура классификации состоит в том, чтобы отнести каждый предъявляемый объект к тому или иному классу. При этом классы могут быть заданы заранее, либо образовываться в ходе обучения, т.е. применение решающих правил.

Одной из важных операций в процессе решения задачи распознавания образов является операция выявления общих характеристик предъявляемых объектов. Отнесение их к одному классу может рассматриваться как обобщение исходных данных. Для этой операции характерны два связанных, противоположно направленных действия — объединение подобных и отделение

отличающихся объектов. Понятие “подобие”, или “сходство” должно быть формализовано.

Пусть C - вектор признаков (измерений) объекта, полученных в результате предварительной обработки. Тогда:

$$G = W * C$$

где G - вектор результата, W - матрица весовых множителей, C - вектор признаков

$$\begin{array}{c} \left| \begin{array}{c} g_1 \\ g_2 \\ g_3 \\ \dots \\ g_i \\ \dots \\ g_n \end{array} \right| = \left| \begin{array}{cccccc} w_{10} & w_{11} & w_{12} & \dots & w_{1p} \\ w_{20} & w_{21} & w_{22} & \dots & w_{2p} \\ w_{30} & w_{31} & w_{32} & \dots & w_{3p} \\ \dots & \dots & \dots & \dots & \dots \\ w_{i0} & w_{i1} & w_{i2} & \dots & w_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ w_{n0} & w_{n1} & w_{n2} & \dots & w_{np} \end{array} \right| \cdot \left| \begin{array}{c} c_1 \\ c_2 \\ c_3 \\ \dots \\ c_i \\ \dots \\ c_n \end{array} \right| \end{array}$$

На основе полученного вектора выбирается решение $d = \max \{g_1, g_2, \dots, g_n\}$

$$d = \begin{cases} x_s \in E_i, \text{ если } G(C_i) = \max \\ x_s \in \bar{E}_i, \text{ если } G(C_i) \neq \max \end{cases}$$

Таким образом, процедура принятия решения есть процедура определения экстремума (максимума) решающей функции G .

В процессе построения решающей функции основная задача состоит в том, чтобы найти весовые коэффициенты для каждого

$$W_i^p = [w_0, w_1, w_2, \dots, w_p].$$

Другой принцип построения решающей функции заключается в ее определении на основе вычисления расстояния между различными классами. Под расстоянием d понимается характеристика, определяющая различие векторов признаков C_I и C_J , относящихся к различным классам E_I и E_J .

Наиболее часто используются следующие меры расстояния [23, 45]:

1. d_1 - евклидово расстояние

$$d_1(\vec{C}_i, \vec{C}_j) = \sqrt{\sum_{k=1}^p |c_{ik} - c_{jk}|^2} \quad (2.10)$$

2. d_2 - расстояние по "Манхэттену" ("матрица городских кварталов")

$$d_2(\vec{C}_i, \vec{C}_j) = \sum_{k=1}^p |c_{ik} - c_{jk}| \quad (2.11)$$

3. d_3 - Чебышевское расстояние

$$d_3(\vec{C}_i, \vec{C}_j) = \max_{k \in \overline{1, p}} |c_{ik} - c_{jk}| \quad (2.12)$$

4. d_4 - расстояние по Канберру

$$d_4(\vec{C}_i, \vec{C}_j) = \sum_{k=1}^p \frac{|c_{ik} - c_{jk}|}{|c_{ik} + c_{jk}|} \quad (2.13)$$

В таком случае выполняется минимизация параметра \mathbf{d} .

Вычисление расстояния может быть положено в основу обучения, т.е. задания новых классов. В качестве значений параметров $c_{jk}; \forall k = \overline{1, p}$ может быть задано усреднение значений параметра по классу E_{ij} т.е.

$$m_{jk} = \frac{1}{S_j} \sum_{C \in E_j} C_{jk}$$

тогда функция сходства

$$Z = \frac{1}{d_j} \quad i = \overline{1, 4}$$

К данному классу методов может быть отнесена и группа дискриминантных методов распознавания [45].

Дискриминантные методы основаны на классификации символов по характерному набору признаков. Это значит, что множество подлежащих обработке символов разбивается на конечное число классов, называемых образами. Объекты каждого класса (образы) характеризуются некоторой совокупностью признаков, каждый из которых принимает одно из определенного множества значений.

Пространство признаков разбивается на непересекающиеся области, каждая из которых соответствует определенному классу символов. Для предъявленного к распознаванию изображения символа вычисляются все признаки, значения которых задают положение некоторой точки в пространстве признаков. После этого определяют, в какую из областей пространства признаков попала данная точка, и исходный символ относят к классу символов, связанному с данной областью.

Описанная процедура распознавания рассматривает каждый образ и каждое конкретное изображение как изолированный объект, пусть довольно сложный, но существующий цельно и независимо от методики решения задачи. Выбор пространства признаков и их значений осуществляется исходя из стремления оптимизировать как разбивку самого пространства признаков на области, так и процедуру

принятия решения. При этом возникают сложности с определением системы устойчивых признаков.

Выбор системы признаков зависит от конкретных особенностей символов, причем система признаков в целом должна удовлетворять следующим условиям:

- система должна содержать минимальное число признаков;
- признаки должны вычисляться с помощью сравнительно простых алгоритмов;
- система признаков должна минимизировать риск ложного распознавания символа.

Система признаков может быть построена на основе совокупности топологических инвариантов, геометрических соотношений и моментов различных порядков.

К топологическим инвариантам может быть отнесено число углов, число пересечений линий в точках, число отдельных различных фрагментов в символе.

В системе геометрических соотношений в качестве признаков используются, например, соотношения сторон выделенного прямоугольника, отношение площади к периметру и т.п.

Достоинством дискриминантного метода является инвариантность к масштабу символа и его пространственному расположению, однако, применительно к задаче распознавания символов такая инвариантность практически не требуется.

К недостаткам данного метода следует отнести достаточно большую трудоемкость подготовки системы признаков и невысокую точность распознавания, что требует обязательного второго этапа распознавания для идентификации символа внутри выявленного класса символов.

2.3.1. Построение решающих функций для двоичных наборов параметров

Пусть имеется два вектора параметров:

$$\vec{C}_i = [c_{1i}, c_{2i}, c_{3i}, \dots, c_{pi}] \quad \vec{C}_j = [c_{1j}, c_{2j}, c_{3j}, \dots, c_{pj}]$$

где $C_{kl} = \{0,1\}$; $\forall k = \overline{1, p}$, $\forall l = \overline{1, n}$

Получаем следующие характеристики:

1. a - число случаев, когда \vec{C}_i и \vec{C}_j обладают одним и тем же признаком

$$a = \sum_{k=1}^p C_{ki} \cdot C_{kj}$$

2. b - число случаев, когда \vec{C}_i и \vec{C}_j не обладают общим признаком

$$b = \sum_{k=1}^p \overline{C}_{ki} \overline{C}_{kj}$$

3. n - число случаев, когда \vec{C}_i не обладает признаком, присущим \vec{C}_j :

$$n = \sum_{k=1}^p \overline{C}_{ki} C_{kj}$$

4. g - число случаев, когда \vec{C}_j обладает признаком, отсутствующим у \vec{C}_i

$$g = \sum_{k=1}^p C_{ki} \overline{C}_{kj}$$

Функция сходства для векторов \vec{C}_i и \vec{C}_j должна обладать следующими свойствами [3]:

- должна быть возрастающей в зависимости от a ;
- убывающей от b ;
- симметричной относительно g и n .

С учетом отмеченных особенностей были предложены следующие функции сходства векторов бинарных признаков [45]:

1. **Рассел и Рао:**

$$S_1(\vec{C}_i; \vec{C}_j) = \frac{a}{a + b + g + h} = \frac{a}{p}$$

2. **Жокар и Нидмен:**

$$S_2(\vec{C}_i; \vec{C}_j) = \frac{a}{p - b}$$

3. **Дайс:**

$$S_3(\vec{C}_i; \vec{C}_j) = \frac{a}{2a + g + h}$$

4. **Сокаль и Мишнер:**

$$S_5(\vec{C}_i; \vec{C}_j) = \frac{a + b}{p}$$

5. **Культинский:**

$$S_6(\vec{C}_i; \vec{C}_j) = \frac{a}{g + h}$$

6. **Юл:**

$$S_7(C_i; C_j) = \frac{ab - gh}{ab + gh}$$

Заметим, что указанные меры пригодны для использования и при выполнении корреляционно-разностного сравнения по битовым картам.

2.4. Контурные методы распознавания

Низкоуровневые и признаковые представления позволяют добиться инвариантности только по отношению к весьма ограниченным классам преобразований. Изменчивость других типов, вызванная, например, сменой сенсора или собственными изменениями объекта, оказывается трудно формализуемой, так как существенно зависит от свойств объектов, образующих наблюдаемую сцену. Важной характеристикой изображения при этом являются пространственные взаимосвязи между его элементами, образующими иерархическую структуру.

Необходимость использования нескольких представлений промежуточного уровня часто иллюстрируют [4], опираясь на аналогию анализа текста (или речи) и изображений. Так же как буквы группируются в слова, слова в предложения, а предложения в текст, пиксели следует группировать в элементы промежуточного представления, которые уже, в свою очередь, образуют конкретные объекты. Подобная аналогия, однако, не говорит, сколько и каких должно быть промежуточных уровней. Могут быть выделены контурные и структурные представления. Как показывает мировой опыт исследователей [1, 23, 24,], именно эти типы представлений оказываются крайне полезными в задачах сопоставления и распознавания изображений в условиях изменчивости.

Контурные представления

Под контуром обычно понимается местоположение локального изменения или резкого перепада яркости на изображении [26]. При этом подразумевается, что такие перепады возникают на границах объектов.

Процедуры построения контурных описаний изображений можно разделить на глобальные и локальные. Для глобальных процедур характерно разбиение изображения на однородные области, на основе которых и строятся контуры, например, как границы этих областей [13, 26] либо как их срединные оси [19].

Локальные процедуры основываются либо на определении цепочек максимумов на градиентном поле изображения (или пересечений нуля второй производной), либо на непосредственной аппроксимации яркостных переходов [26] в рамках функциональных или стохастических моделей изображений. Контуры могут трактоваться также как цепочки точек интереса, обладающих

экстремальными значениями некоторого признака края [45], что, однако, не влияет на сущность методов их обнаружения.

Построение градиентного поля полутонового изображения осуществляется с помощью операторов, выполняющих дискретное дифференцирование. Классическими операторами, служащими для этой цели, являются операторы Робертса, Превитта, Собела и другие [13, 16, 26].

Могут также использоваться и маски больших размеров. Поскольку таким операторам присуще свойство подавления шума, увеличение размера маски приводит к большей робастности процедур извлечения контуров, однако и увеличивает вероятность обнаружения ложных контуров [26]. Помимо получения более гладких контуров, удобных для дальнейшего анализа, привлечение масок разных размеров обосновывается иерархичностью организации физического мира [19].

Были предложены различные методы для формирования контуров на разных уровнях разрешения. Так, например, в работе [19] обосновано использование фильтра $\nabla^2 G$, где ∇^2 оператор Лапласа и

$$G = \exp\left[-\frac{x^2 + y^2}{2\pi\sigma^2}\right].$$

определяет Гауссову функцию. Более того, выглядит правдоподобным, что именно эту функцию воспроизводят X-клетки сетчатки и наружного коленчатого тела зрительного аппарата человека [19].

Широко применяются и другие подходы к обнаружению контуров, такие как, например, предложенные Линденбергом [82] и другие.

Таким образом, построение градиентного поля изображения основывается на какой-либо математической модели. Затем осуществляется выделение контуров как цепочек максимумов на этом поле. Однако возникает проблема обоснования формализации понятия контура (о существовании этой проблемы свидетельствует разнообразие подходов к обнаружению контуров).

Как правило, приводятся следующие аргументы в пользу привлечения контуров [43]:

- \square контур является концентратором информации в изображении;
- контур полностью характеризует форму объектов на изображении;

- контуры объекта, в отличие от его остальных точек, устойчивы на изображениях, полученных в разное время, разных ракурсах и при смене датчика;

- контурные точки составляют незначительную часть всех точек изображения, поэтому работа с ними позволяет резко сократить объем вычислений.

В несколько других терминах аргументация использования контуров заключается в том, что нарушения непрерывности в физическом пространстве (поверхности) порождают также и перепады яркости на изображении [19]. А значит, последние могут быть использованы для получения информации о реальных поверхностях.

Контурные методы распознавания изображений

Использование контуров широко распространено в задачах совмещения пары изображений или изображения с векторной моделью (например, картой местности или чертежом детали), описание формы объектов или областей по их контурам, например, с помощью методов математической морфологии [13], для решения задачи стереоскопического зрения. Контурные представления также служат основой для построения структурных описаний изображений.

При распознавании изображений в рамках контурного подхода обычно используется преобразование расстояния [4, 45], привлекаемое для вычисления меры Хаусдорфа [98]. При выполнении преобразования расстояния для каждой точки изображения определяется минимальное расстояние до множества точек контуров. Далее на карту расстояний накладываются контуры другого изображения (при некоторой гипотезе сопоставления), что позволяет вычислить суммарное расстояние между двумя наборами контуров. При этом, как правило, необходимо осуществлять перебор значений параметров взаимного преобразования изображений, поэтому пространство поиска должно быть не слишком большим (либо число параметров должно быть мало, либо на них должны быть наложены достаточно сильные ограничения).

Итак, хотя контурные методы распознавания являются весьма робастными [43], в них оказывается затруднительным поиск пространственных преобразований сложного типа, а также распознавание объектов при наличии большого числа ложных контуров.

Заметим, что в результате выделения контуров и при использовании для описания выделенного контура, например, цепного кода Фримена, получается описание в виде списка. Элементами такого списка будут трехбитные коды Фримена [45]. В этом случае задачу распознавания контуров можно свести к задаче

вычисления меры близости (расстояния) между двумя описаниями-списками кодов.

2.4.1. Определение расстояния между описаниями объектов

В ряде случаев распознаваемые объекты могут быть заданы формализованным описанием на каком-либо языке. Такие описания могут рассматриваться как некоторые упорядоченные списки символов, а задача распознавания объектов может быть сведена к задаче определения расстояния между списками [45].

Определение расстояния между двумя списками состоит в учете последовательности составляющих в каждом списке. При этом учитывается структура последовательности.

Каждый список состоит из символов, входящих в один алфавит. Они могут быть представлены в виде типографских знаков (например, букв), или других элементов, образующих конкретный текст.

Пусть заданы два списка:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\} \text{ и } X_j = \{x_{j1}, x_{j2}, \dots, x_{jm}\}$$

В общем случае n и m могут быть не равны друг другу. Задача заключается в том, чтобы определить такую функцию, которая выражала бы степень сходства этих 2 списков.

Из рассмотрения 2 последовательностей видно, что одна из них может быть преобразована в другую. Например, чтобы перейти от ряда ABC к ряду ABD , следует сохранить 1-й и 2-й символы, заменить 3-й и добавить 4-й. Введя пустой символ λ , можно эти 3 варианта преобразований записать в виде 3 операций:

SUBstitution (подстановка) $x_i \rightarrow x_j$ *SUB*(x_i, x_j);

DEStuction (уничтожение) $x_i \rightarrow \lambda$ *DES*(x, λ);

CREation (создание) $\lambda \rightarrow x_j$ *CRE*(λ, x);

Каждому преобразованию соответствует своя цена c (...), а именно:

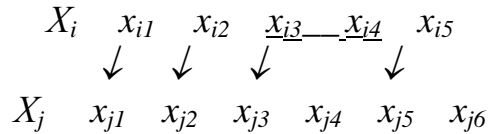
$c(x_i, x_j)$ для *SUB*;

$c(x, \lambda)$ для *DES*;

$c(\lambda, x)$ для *CRE*;

Для оценки расстояния между двумя списками вводят понятие полной цены последовательности преобразований как наименьшей из всех возможных цен, которые нужно “уплатить” за переход от исходного списка к конечному. Расстояние $\delta(X_i, X_j)$, соответствующее полной цене, является минимальным. Это означает, что если $\delta(X_i, X_j) = C(X_i, X_j)$, то переход от X_1 к X_2 происходит без получения промежуточных последовательностей.

Схема преобразования носит название пути, обычно изображаемого стрелками. Так, путь преобразований от X_i к X_j , следующий:



Его можно записать в виде ряда операций:

$SUB(x_{i1}, x_{j1}); SUB(x_{i2}, x_{j2}); DES(x_{i3}, \lambda); SUB(x_{i4}, x_{j3}); CRE(\lambda, x_{j4});$
 $SUB(x_{i5}, x_{j5}); CRE(\lambda, x_{j6}).$

Каждой операции приписывают определенную цену, и расстояние между списками соответствует минимуму цены. Рассмотрим еще 2 списка:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$$

$$X_j = \{x_{j1}, x_{j2}, \dots, x_{jm}\}.$$

Положим $X_i(l) = \{x_{i1}, x_{i2}, \dots, x_{il}\}$ и $X_j(k) = \{x_{j1}, x_{j2}, \dots, x_{jk}\}$, тогда, например,

$$X_{i(5)} = \{x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}\} \text{ и } X_{j(3)} = \{x_{j1}, x_{j2}, x_{j3}\}.$$

Положим $D(l, k) = \delta[X_i(l), X_j(k)]$. Для перехода от $D(l-1, k-1)$ к $D(l, k)$ имеются три возможности:

- путем подстановки $X_i(l)$ и $X_j(k)$, цена которой соответствует цене подстановке $SUB(x_{il}, x_{jk})$ последних элементов в каждом списке;
- путем создания последнего элемента в списке j , за что придется уплатить цену $CRE(\lambda, x_{jk})$;
- наконец, если рассматривать пару $X_i(l-1)$ и $X_j(k)$, то цена будет $DES(x_{il}, \lambda)$, что соответствует стоимости уничтожения последнего элемента в списке i .

Кратко эту процедуру можно записать так:

$$D(l, k) = \text{Min} \begin{cases} D(l-1, k-1) + c(x_{il}, x_{jk}) \\ D(l, k-1) + c(\lambda, x_{jk}) \\ D(l-1, k) + c(x_{il}, \lambda) \end{cases}$$

и минимальная цена соответствует оптимальному пути. Запишем это в виде алгоритма:

1. $D(0,0) = 0$ (начальное значение)

2. Для $l:=1$ до n

$$\text{выполнять } D(l,0) := D(l-1,0) + c(x_{il}, \lambda)$$

3. Для $k:=1$ до m

$$\text{выполнять } D(0,k) := D(0,k-1) + c(\lambda, x_{jk})$$

Для $l:=1$ до n

для $k:=1$ до m

$$\text{выполнять } d1 := D(l-1, k-1) + c(x_{il}, x_{jk})$$

$$d2 := D(l, k-1) + c(\lambda, x_{jk})$$

$$d3 := D(l-1, k) + c(x_{il}, \lambda)$$

$$D(l, k) = \text{Min}\{d1, d2, d3\}$$

5. Минимальное расстояние $\delta(X_i, X_j) = D(n, m)$

2.5. Вероятностный подход к распознаванию Моделей на основе случайных полей

Продуктивной концепцией является представление изображений как выборочных функций (реализаций) случайного процесса. Такие модели являются более общими, чем функциональные, так как представляют изображения некоторыми случайными функциями, по отношению к которым детерминированные функции являются частным подклассом. Однако зачастую модели на основе случайных полей описывают статистические свойства изображений, не отражая при этом их пространственной структуры. Стремление включить в стохастическую модель информацию о пространственных положениях элементов изображения приводит, как правило, к использованию различных моделей случайных полей [92], однако, редко используемых в задачах распознавания.

Стохастические модели изображений берут свое начало из теории цифровой обработки сигналов. Исходно они предназначались для решения задач подавления шума, помехоустойчивого кодирования и сжатия, в частности, в целях передачи изображений по каналам связи [26]. Сейчас стохастические модели положены в основу многих графических форматов, которые используются для хранения изображений в информационно-телекоммуникационных системах, а также широко используются для описания текстуры [75], реставрации изображений [13], для декомпозиции изображений на области [13, 54] и других задач [97].

В рамках порождающих стохастических моделей изображений предполагается, что имеется набор параметров $\chi = (\chi_1, \dots, \chi_N)$ с заданной плотностью распределения априорных вероятностей $P(\chi)$, а также функция построения изображения, задающая распределение вероятностей по изображениям при данных параметрах: $P(f | \chi)$. Моделью конкретного изображения будет являться набор значений этих параметров.

Сопоставление изображений в рамках стохастических моделей

Использование стохастических моделей изображений позволяет использовать в качестве меры сходства вместо корреляции взаимную информацию, вычисляемую через энтропию H

$$i(f_1(\mathbf{x}), f_2(T(\mathbf{x}))) = H(f_1(\mathbf{x})) + H(f_2(T(\mathbf{x}))) - H(f_1(\mathbf{x}), f_2(T(\mathbf{x}))),$$

где T – некоторое пространственное преобразование, а вычисления энтропии некоторой случайной величины и совместной

энтропии двух случайных величин осуществляются как

$$H(f) = -\sum_f p(f) \log_2 p(f) \text{ и } H(f_1, f_2) = -\sum_{f_1} \sum_{f_2} p(f_1, f_2) \log_2 p(f_1, f_2).$$

Величина взаимной информации является инвариантной к произвольным глобальным (не зависящим от координат) преобразованиям яркости изображений, а не только к линейным преобразованиям, как это имеет место в случае корреляции, что делает эту меру сходства существенно более универсальной. К сожалению, при этом оказывается необходимым вести поиск в пространстве параметров преобразования T , что уже для преобразования подобия может оказаться проблематичным в силу увеличения размерности пространства параметров.

Поскольку стохастические модели изображений характеризуются определенными параметрами, оказывается возможным отделить случайную составляющую изображений (конкретную реализацию случайного процесса) от регулярной составляющей (параметров модели). Возможно выполнение сопоставления изображения с эталоном по этим параметрам. Однако описание изображения через параметры модели следует рассматривать в качестве признакового представления изображений.

Таким образом, функциональные и стохастические модели преимущественно используются на предварительном этапе обработки изображений и построения описаний последующих уровней (признакового, контурного, структурного), либо для задач распознавания изображений, обладающих малой изменчивостью.

Если решающая функция имеет статистический характер, то можно воспользоваться вероятностным подходом к распознаванию [23, 32,45].

Пусть E_i - некоторый класс из множества V рассматриваемых классов, который появляется с вероятностью $P(E_i)$, а \vec{C} - вектор параметров. Тогда условная вероятность получения вектора \vec{C} при условии, что известен класс E_i будет $P(\vec{C} | E_i)$

По закону Байеса, условная вероятность класса E_i при известном \vec{C} может быть определена так:

$$P(E_i | \vec{C}) = \frac{P(E_i) \cdot P(\vec{C} | E_i)}{P(\vec{C})}$$

Вероятность $P(\vec{C} | E_i)$ носит название функции правдоподобия. Её можно вычислить заранее или получить опытным путем (сбором статистики).

В результате полагают, что объект принадлежит к классу E_i при данном векторе параметров, если $P(E_i|\vec{C}) = \max (\forall i = \overline{1, n})$

Основным недостатком подобного метода является то, что нужно собрать данные, т.е. знать достаточно длинную предысторию.

2.6. Структурные методы распознавания

Если исходные изображения можно считать (как правило) двумерным сигналом, контуры на изображениях – одномерным сигналом, то следующие по уровню абстракции промежуточные символичные представления изображений являются безразмерными (или, условно, с размерностью, равной нулю). Однако способы получения таких представлений могут быть разными.

2.6.1. Непроизводные структурные элементы

В качестве основы структурных представлений изображений могут быть использованы точки интереса, строящиеся в рамках признакового подхода [4]. При этом, как правило, используются более сложные детекторы, чем детекторы края. Могут привлекаться детекторы линий, полос, пятен или углов, находящихся в соединениях отрезков вида “Г”, “Т”, “У” и “Х” [44, 45]. Используются также детекторы признаков, предназначенные для узких предметных областей, например, для обнаружения лиц [10, 35, 66]. Исследуется и противоположная возможность, заключающаяся в автоматическом построении детекторов признаков на основе содержащейся в изображении информации [78].

Более последовательным для построения описаний изображений данного уровня абстракции является использование представлений предыдущего уровня, то есть контурных представлений. Построение структурных элементов на основе контуров может выполняться путем сегментации контуров, их аппроксимации или обнаружения точек максимальной кривизны. При этом контурные структурные элементы являются менее разнообразными, чем элементы, построенные на основе детекторов признаков, и обычно включают отрезки прямых линий, дуги окружностей или эллипсы, углы и т.д. [23,26,44,45]

Существует также и промежуточный подход, при котором выделяются геометрические элементы на основе краевых точек, не объединенных в связанные контуры [13, 45]. Этот подход более распространен в промышленных приложениях технического зрения и фотограмметрии, поскольку краевые точки могут быть локализованы с субпиксельной точностью, что позволяет выполнять более точные измерения параметров геометрических элементов, например, с помощью метода наименьших квадратов [46]. При этом, однако,

точки, принадлежащие разным элементам, не разделены. Поэтому здесь применяются специфические методы, которые одновременно решают проблему группирования и регрессии. Среди методов этого вида наибольшей популярностью пользуется преобразование Хафа [13].

2.6.2. Составные структурные элементы

Дальнейшее развитие структурного подхода заключается в формировании на основе производных элементов составных структурных элементов. Типы составных структурных элементов могут быть как проблемно-зависимыми, так и достаточно общими. В качестве первых могут служить, например, такие элементы, составленные из отрезков прямых линий, как проекции прямоугольных параллелепипедов на плоскость (в более простом случае ищутся также параллельные линии, прямоугольники и параллелограммы), нахождение которых необходимо, например, для обнаружения зданий и дорог на аэрокосмических изображениях [4, 23, 47]. Последнее важно в гео-информационных системах для извлечения изображения по содержанию.

Существуют различные подходы к группированию производных структурных элементов с целью образования более сложных структурных элементов, вид которых не зависит от предметной области. К таким элементам относятся, например, элементы вида "L", "U" и другие соединения, пары параллельных прямых линий, пятна, составленные из пятен меньших размеров, и т.д. [19]. Однако все подходы к группированию в том или ином виде опираются на подобие структурных элементов. На основе подобия составные структурные элементы могут дальше группироваться, формируя более крупномасштабные элементы изображения.

Попытки строгого обоснования различных типов составных структурных элементов практически отсутствуют. Имеется лишь общая концепция Д. Марра [19] об иерархической пространственной организации реального мира, которую также должны отражать иерархические структурные представления изображений.

Структурные методы распознавания изображений

При структурном сопоставлении изображений, как правило, производится поиск соответствий между структурными элементами с учетом их взаимного расположения, типа, размеров, взаимной ориентации и т.д. Поиск может осуществляться перебором возможных комбинаций соответствий [23], методами эволюционных вычислений [79], нейронными сетями [42], методами итеративной оптимизации [4], с помощью преобразования Хафа [13] и т.д. Иными словами, в структурных методах вместо поиска в пространстве параметров пространственного преобразования производится поиск в пространстве вариантов соответствий структурных элементов.

К сожалению, в большинстве случаев привлекаемые структурные представления не полностью инвариантны по отношению условиям формирования изображений. В связи с этим, на изображениях, полученных в разных условиях, структурные элементы выделяются по-разному, что уменьшает вероятность корректного сопоставления и снижает его точность. Тем не менее, структурные методы принято считать одними из наиболее робастных [45]. В частности, с помощью структурных методов могут сопоставляться аэрокосмические изображения, полученные с помощью сенсоров различных типов и в разные сезоны [23], что является существенным при построении геоинформационных систем. Также в рамках данного подхода могут сопоставляться [24] изображения с векторной моделью или наброском, сделанным от руки, что может быть использовано для извлечения изображений из баз данных по содержанию.

В то же время, из-за отсутствия общей и ясной теории построения структурных представлений создание системы структурного распознавания изображений может быть сопряжено с определенными трудностями. В ряде случаев из-за потери большого количества информации в процессе построения структурных описаний происходит чрезмерное обобщение, в результате чего могут отождествляться изображения с разным содержанием [24]. В зависимости от задачи это может быть как положительным, так и отрицательным свойством. Кроме того, структурными методами плохо распознаются малоразмерные (менее 50x50 пикселей) и слабоструктурированные (например, дымы) объекты.

Данный класс методов распознавания принципиально отличен от рассмотренных ранее [45].

Он основан на анализе структуры объекта и связей между объектами. Такой подход предполагает, что структура (или "конструкция") всех подлежащих распознаванию объектов подчиняется вполне определенным, заранее известным правилам. Данный математический аппарат основывается на аналогии между структурой изображения и структурой фразы в естественном или искусственном языке. Подобно тому, как в языке фраза строится из слов, а слова из букв, можно построить описание (грамматику) изображения, выделив простейшие фрагменты и связи между ними. Если грамматика построена успешно, то возникает возможность анализировать изображения, пользуясь синтаксическими правилами этой грамматики, как при синтаксическом разборе предложений естественного языка. Множество грамматик, обладающих схожими признаками, объединяют в грамматические классы.

Такой подход предполагает, что конструкция, или структура, всех подлежащих распознаванию объектов подчиняется вполне

определенным, заранее известным правилам. Распознавание объекта при таком методе выполняется в два этапа:

- этап составления описания объекта или кодирования;
- этап синтаксического разбора (анализа) составленного описания.

На первом этапе разрабатываются те правила, которым подчиняются конструкции; иначе говоря, создается своего рода грамматика.

В качестве признаков при составлении описания используются характерные точки объекта: начальные точки линий, точки пересечения линий; точки перегиба линий и т.д. В результате описание объекта имеет вид списка (вектора описаний примитивов), который содержит информацию как о примитивах, входящих в состав объекта, так и о их взаимном расположении.

Когда грамматика определена, наступает второй этап, состоящий в принятии решения о том, принадлежит ли рассматриваемый объект к множеству всех объектов, порождаемых этой грамматикой.

Когда грамматика определена, то выполняется второй этап. На нем принимают решение о том, принадлежит ли данный объект к множеству объектов, порождаемых данной грамматикой.

При синтаксическом разборе (анализе) составленного описания проверяется, соответствует ли данное описание какому-либо из описаний объектов-эталонов, подготовленных заранее и хранящихся в базе данных. Такой поиск может выполняться путем попарного сравнения списков. Для сокращения времени сравнения поиск может выполняться путем предварительной сортировки объектов на группы по характерным примитивам и поиск выполнять по двоичному дереву.

Грамматикой называется конечное множество правил, определяющих способ образования цепочек описаний (предложений) из множества символов языка. Формально любая грамматика может быть описана как множество G :

$$G = (\{V_N\}, \{V_T\}, \{P\}, S) \quad (2.14)$$

где $\{V_T\}$ - множество терминальных признаков объекта (терминальный словарь); $\{V_N\}$ - множество нетерминальных признаков (нетерминальный словарь); $\{P\}$ - набор правил подстановки, определяющий способы формирования того или иного объекта или его частей из терминальных и нетерминальных признаков; S - начальный (корневой) признак грамматики или аксиома.

Например, в лингвистике словари V_N и V_T создаются с помощью символов, в качестве которых используются буквы алфавита (например, латинского, греческого и т.д.). В роли алфавита может выступать и совокупность каких-либо графических символов. Процесс

синтаксического распознавания может быть представлен следующим образом (рис. 2.4):

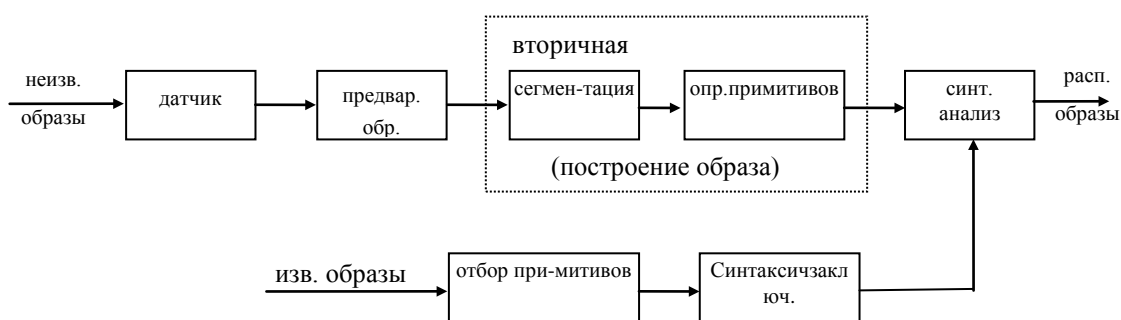


Рис.2.4. Основные этапы синтаксического распознавания

Проблема построения эффективной языковой системы для описания изображений может решаться двумя путями [38]. Один из них состоит в том, что алфавит и грамматика выбираются человеком. Другой путь предполагает автоматическое формирование языка. В этом случае считается, что машина, наблюдая изображения некоторого класса, сама должна сформировать соответствующий алфавит и набор правил описания изображений в этом алфавите.

Многие задачи обработки изображений вообще выходят за рамки метода классификации по совокупности признаков. К числу таких задач относится обработка изображений, не являющихся неделимыми с общечеловеческой точки зрения, например, буква или цифра.

Анализ фотографий с искусственных спутников Земли земной поверхности с целью выявления природных ресурсов, анализ принципиальной схемы электронного устройства при разработке фотошаблонов для плат печатного монтажа, являясь, по сути, задачами распознавания, не могут быть сведены к трем простейшим ответам "да", "нет", "не знаю", а требуют глубокого проникновения в структуру изображения - сцены. Задачи такого типа определяются как "анализ сцен".

При решении таких задач часто требуется дать ответ на вопрос: находится ли определенный объект в изображении - сцене, находятся ли выделенные объекты сцены в заданном положении, например один выше другого, слева от него, за другим и т.д. Может потребоваться и описание сцены. Для решения такого рода задач необходимо иметь развитый аппарат описания сцены, ориентированный на иерархическую структуру ее объектов, опирающийся на систему производных (простейших) элементов и отношений между ними.

Развитие такого аппарата основывается на аналогии между структурой изображения и структурой фразы в естественном или искусственном языке. Подобно тому, как, опираясь на правила грамматики, мы строим фразы из слов, а слова из букв, можно построить изображение (сцену) из объектов, опираясь на отношения между объектами, а объекты состоят из соединяемых по определенным правилам легко выделяемых простейших фрагментов - производных элементов. Если удастся найти такую грамматику изображений, то возникает возможность анализировать изображения, пользуясь синтаксическими правилами этой грамматики как при синтаксическом разборе предложений естественного языка. В этом случае говорят о лингвистическом подходе к распознаванию, иногда его называют структурным, а соответствующие формальные средства - языком описаний изображений.

Лингвистический аппарат описания изображений можно применять не только для сцен или изображений со сложной иерархической структурой. При соответствующем подборе грамматики он может применяться и для анализа и распознавания таких "простых" образов, как, скажем, буквы или цифры.

Попытки применить методы математической лингвистики к задаче анализа изображений приводят к необходимости решить ряд проблем, связанных с отображением двумерной структуры изображения на одномерные цепочки формального языка.

1. В качестве производных элементов выбирают сложные элементы, например штрихи (отрезки прямых) — горизонтальные, вертикальные, наклонные, крючки, кривые и т.д. Выявление этих примитивов может оказаться непростой задачей, поскольку довольно часто нет четкой границы между, скажем, вертикальной и наклонной линией (например, в рукописных символах). Еще сложнее отличить кривые линии от наклонных и т.д., особенно учитывая ограниченные размеры растра. Использование в качестве производных элементов определенных областей символа равносильно решению задачи его центровки на растре, а сделать это не просто.

2. Производные элементы воспринимаются со значениями их признаков, например, для отрезков рассматривается признак "толщина", или "ширина". Толщина линий, даже относительная, существенно зависит от условий написания.

3. При выборе как примитивов, так и отношений между ними (правил композиции) совершенно не учитывают вопросы технической реализации распознающей системы. На первый взгляд может показаться, что эти вопросы не имеют существенного значения, поскольку методы ориентированы на использование универсальной ЭВМ. Однако необходимо помнить, что практические системы распознавания букв и цифр предназначены для скоростного ввода

информации в ЭВМ, и время распознавания каждого символа является для них чрезвычайно существенным технико-экономическим показателем.

Серьезным недостатком синтаксического метода является неустойчивость распознавания при наличии случайного шума.

2.7. Методы распознавания, основанные на знаниях

Как отмечалось выше, в качестве конечного представления изображений в задачах их распознавания выступает некоторая система представления знаний. Более того, многие авторы указывают на необходимость использования знаний и в самом процессе анализа изображений [83, 84]. Аргументация здесь используется примерно следующая: если мы издали видим темное пятно над столом, то мы можем догадаться, что это телефон, хотя информации в самом изображении телефона для этого недостаточно, то есть мы используем для анализа изображения высокоуровневую информацию. В связи с этим полагается, что и весь процесс обработки зрительной информации должен вестись под управлением знаний (подход сверху вниз). В качестве основной аргументации [4] в пользу этого выступает утверждение о недостоверности результатов, полученных в подходах, ведомых данными (подходах снизу вверх), и о проблеме комбинаторного взрыва количества возможных различных интерпретаций.

В классическом распознавании изображений знания описываются в терминах тех представлений, которые используются в самих алгоритмах распознавания. Распознавание может вестись на основе каких-либо низкоуровневых представлений, тогда и объект описывается соответствующим образом, например, в виде его инвариантных признаков [23, 45], могут храниться в виде контурных [43] или структурных [44] описаний, если привлекаются более абстрактные представления.

Таким образом, распознавание часто сводится к получению описания изображения и его сравнению (вычисление некоторой меры сходства) с описанием объекта, поиск которого производится.

Если в классическом распознавании изображений знания представляются неявно, то в последние десятилетия появилась тенденция явного представления знаний в системах автоматической интерпретации изображений. Здесь выделяют три типа знаний [4]: перцепционное, семантическое и функциональное знание. Перцепционное знание позволяет интерпретировать изображения в терминах линий, областей и т.д., в то время как семантическое знание описывает определенные абстрактные понятия, такие как форма или конкретные объекты и отношения между ними.

Функциональное знание предназначено для регулирования процесса интерпретации изображения в зависимости от предметной области и поставленной цели.

Существуют следующие подходы к представлению знаний в системах анализа изображений [4]: семантические сети, объектно-ориентированные представления и фреймы, продукционные системы, мультиагентный подход, представления, основанные на логике предикатов, и некоторые другие. Часто для записи знаний различного типа в одной системе может привлекаться несколько представлений. Все эти представления заимствованы из теории экспертных систем.

Как правило, системы, основанные на знаниях, априорно имеют высокоуровневое описание сцены [16]. Задача заключается в привязке этого высокоуровневого описания к изображению, а вовсе не в построении самого описания. Высокоуровневое описание сцены задается не полностью: могут быть неточно известны положения объектов, какие-то объекты или их части в априорном описании могут быть пропущены из-за неполноты этого описания или могут быть указаны лишние, если часть объекта закрыта. Однако выбор всегда осуществляется из малого числа альтернативных описаний сцены, каждое из которых выдвигается в качестве гипотезы и проверяется на соответствие с изображением. Это позволяет избежать комбинаторного взрыва числа возможных описаний, характерного для подхода снизу вверх, однако методы, основанные на знаниях, способны работать лишь в сильно ограниченных предметных областях.

Более того, при разработке систем, основанных на знаниях, также возникает потребность решать проблему построения промежуточных описаний изображений. Например, в работах [54,70] используется одна и та же система AIDA (в ней семантические знания представляются с помощью семантических сетей, а функциональные – с помощью продукционных правил) для двух разных задач: восстановления трехмерной формы близких объектов и выбора опорных точек на аэрокосмических изображениях.

Однако в каждом из применений этой системы оказывается необходимым решать свои задачи более низкого уровня: сегментация по яркостному изображению и карте глубины в первом случае и выделение узких полос на изображении (параллельных контуров на градиентном изображении) во втором случае. В системах, подобных системе AIDA, также используется ряд концептов (например, полигон) и взаимосвязей между ними (например, понятие перпендикулярности), которые неявно закладываются в алгоритмы первоначального анализа изображений. Лишь иногда эти понятия выделяются в явном виде в качестве перцепционных знаний. Поскольку полнота системы низкоуровневых понятий не исследуется,

далеко не любые высокоуровневые знания могут быть представлены в рамках выбранного формализма (например, семантических сетей).

Значит, при применении подобных систем к конкретным задачам значительная часть работы переносится на разработчика, который выбирает низкоуровневые признаки и реализует эвристические процедуры их выделения. При этом универсальный набор необходимых низкоуровневых признаков не обсуждается. Таким образом, прежде чем рассматривать возможность создания системы распознавания, способной строить семантическое описание произвольной сцены в рамках некоторого представления знаний, необходимо решить проблему построения промежуточных представлений.

Представления, основанные на знаниях, используются и для распознавания сцен. Эти методы весьма схожи с методами структурного сопоставления, где структурные элементы заменяются метками, соответствующими реальным объектам на изображениях. В отличие от структурных элементов, метки являются более уникальными и легче отождествляются друг с другом. Как уже отмечалось, существующие представления, основанные на знаниях, полностью зависят от предметной области и для переноса на другую область требуют существенных затрат. Однако подход на основе знаний может быть эффективно использован в случае наличия текстовых аннотаций к распознаваемым изображениям [87].

2.8. Критерии эффективности распознавания изображений

Эвристические методы

Критерий, на основе которого определяется степень соответствия двух изображений при их взаимном преобразовании, является существенным аспектом методов распознавания. В то же время во многих методах используются критерии, применимость которых строго не обосновывается, либо обосновывается в рамках очень сильных ограничений. К подобным критериям относятся, к примеру, коэффициент корреляции для значений яркостей [13, 25], мера Хаусдорфа для контуров [98, 100] или среднеквадратичное отклонение для структурных элементов или точек интереса [23]. Хотя все эти критерии и имеют некоторую математическую основу, их применение к реальным изображениям во многом эвристично. В частности, это связано с тем, что использование подобных критериев для точек интереса, контуров или структурных элементов осуществляется безотносительно к способу получения тех или иных сопоставляемых элементов изображения. Кроме этого, каждый из указанных критериев привлекается для конкретного типа представлений изображений. В работе [23] показано, что эти

критерии могут быть получены как частные следствия более общих критериев, таких как апостериорная вероятность или количество информации, при наложении ряда упрощающих ограничений. Эти ограничения (часто не указываемые при описании того или иного критерия и присутствующие в них лишь неявно) необходимо учитывать на практике.

Помимо частных критериев, которые все же имеют какое-то обоснование, часто привлекаются *ad hoc* критерии [4]), подобранные на основе опыта разработчиков и их здравого смысла для конкретной задачи. Такие критерии обладают большим разнообразием и низкой достоверностью. В связи с этим, эвристические методы рекомендуется применять только на начальном этапе проектирования систем анализа изображений.

Байесовские методы

Байесовский подход к анализу изображений является одним из наиболее разработанных общих и строго обоснованных подходов к заданию критерия качества решения. Данный подход может применяться как для распознавания изображений на основе различных типов представлений [45], так и для построения описаний изображений в рамках этих представлений, а также решения ряда других задач анализа изображений [13].

В рамках данного подхода решение выбирается на основе его апостериорной вероятности, вычисляемой по правилу Байеса:

$$P(h|f) = \frac{P(h)P(f|h)}{P(f)},$$

где $P(h|f)$ - апостериорная вероятность решения (гипотезы) h при данном изображении f , $P(f|h)$ - функция правдоподобия данных (изображения) f в рамках решения h , $P(h)$ - априорная вероятность решения h , $P(f)$ - априорная вероятность появления изображения f , не влияющая на выбор решения.

В качестве решения h может выступать описание любого уровня (признаковое, контурное, структурное), в том числе и результат распознавания изображения. К примеру, в рамках низкоуровневых методов предполагается, что имеется набор $\chi = (\chi_1, \dots, \chi_N)$ с заданной плотностью распределения априорных вероятностей $P(\chi)$, а также функция построения изображения, задающая распределение вероятностей по изображениям при данных скрытых переменных $P(f|\chi)$; где f – изображение. Описанием конкретного изображения будет являться набор значений этих скрытых переменных.

Одной из простейших является модель, в которой интенсивности отдельных пикселей считаются распределенными по

одному и тому же нормальному закону, характеризующемуся двумя параметрами: a (средним) и σ (дисперсией):

$$p(f | (a, \sigma)) = \prod_{(x,y)} p(f(x,y)) = \prod_{(x,y)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(f(x,y)-a)^2}{2\sigma^2}\right].$$

Эта модель, очевидно, очень упрощенная, так как она не учитывает пространственные зависимости, которые как раз и характеризуют изображения. Однако в силу своей простоты она является удобной стартовой точкой при построении более сложных статистических моделей. Более сложные модели могут включать учет статистик высоких порядков, ограничения на пространственный спектр или некоторые общие предположения о свойствах изображений, такие как инвариантность к масштабу или стационарность [45].

Однако повышению адекватности стохастических моделей препятствуют два обстоятельства: трудность явного задания распределения вероятностей $P(f|h)$ для высокоуровневых представлений изображений и неизвестность априорного распределения вероятностей $P(h)$, которое существенным образом сказывается на выборе решения, но при этом в рамках байесовского подхода не может быть строго определено.

При математической корректности и широкой распространенности байесовского подхода он обладает рядом методологических недостатков, связанных с выбором распределения априорных вероятностей и заданием плотности распределения условных вероятностей на множестве изображений, что накладывает ограничения на применение данного подхода совместно с некоторыми типами представления изображений.

Энтропийные методы

В рамках энтропийного подхода к анализу изображений основной величиной, на основе которой осуществляется выбор решения, является энтропия, введенная в шенноновской теории информации. В этой теории количество информации, содержащееся в некотором сообщении, определяется как

$$I(x) = -\log_2 P(x),$$

где x – значение некоторой случайной величины, $P(x)$ – вероятность принятия случайной величиной данного значения, $I(x)$ – количество собственной информации в сообщении, содержащем значение x . Математическое ожидание $E_X[I(X)]$ количества информации $I(x)$ называется энтропией $H(x)$ данной случайной величины x

$$H(X) = E_X[I(X)] = \sum_{x \in X} I(x)P(x) = - \sum_{x \in X} P(x) \log_2 P(x).$$

Энтропия значений, хранящихся в пикселях изображения, $H(f(x, y) | h)$, вычисленная при некоторой гипотезе h , выступает в качестве критерия выбора решения. Согласно принципу минимума энтропии [17], лучшим является решение, для которого достигается минимум энтропии.

Данный принцип может быть выведен из правила Байеса. В результате логарифмирования можно получить

$$I(h | f) = I(f | h) + I(h) - I(f).$$

Если пренебречь значением $I(h)$, то выбор решения будет осуществляться только на основе величины $I(f | h)$. Для изображений эта величина имеет вид энтропии. К примеру, если точки изображения рассматриваются как независимые, то вероятность для всего изображения представляется как произведение вероятностей для каждой его точки:

$$p(f | h) = \prod_{(x, y)} p(f(x, y) | h),$$

Откуда

$$I(f | h) = - \sum_{(x, y)} \log_2 p(f(x, y) | h) = - \sum_z p(f(x, y) = z | h) \log_2 p(f(x, y) = z | h) = H(f(x, y) | h).$$

Поскольку в шенноновской теории количество информации вычисляется через вероятности, неудивительно, что энтропийные методы анализа изображений наследуют все те трудности, с которыми сталкиваются байесовские методы. В частности, вычислить количество информации $I(h)$, содержащейся в гипотезе h , невозможно без знания априорной вероятности $P(h)$. В принципе минимума энтропии эта трудность игнорируется.

Данный принцип, однако, дополняется важным принципом максимума энтропии [105]. Согласно этому принципу, если какие-то два статистических распределения одинаково хорошо описывают наблюдаемые частоты значений некоторой случайной величины, то следует выбрать распределение, обладающее максимальной энтропией. Принцип максимума энтропии может служить руководством для задания априорных распределений вероятностей.

Плотность распределения случайной величины выступает в роли статистической модели источника информации. Выбор плотности распределения с максимальной энтропией – это выбор модели, в которую не привнесено информации, помимо той, которая содержится в исходных данных. Это часто отождествляют с выбором наименее детальной или наиболее простой модели [105]. Сочетание принципов минимума и максимума энтропии приводит к подходу минимакса энтропии [105], в котором выбор модели осуществляется

как поиск компромисса между ее точностью и простотой, выраженных в форме энтропии.

Таким образом, энтропийные методы имеют много общего с байесовскими методами. В них остается нерешенной проблема выбора формы распределения условных вероятностей и лишь частично решена проблема выбора априорных вероятностей решений. Энтропийные методы нашли широкое применение в задачах совмещения изображений, выявления изменений в серии изображений, извлечении признаков и построении контуров, реставрации изображений, распознавании объектов [4].

Теоретико-информационные методы

Теоретико-информационный подход к анализу изображений основывается на принципе минимальной длины описания. Данный подход базируется не на шенноновской, а на алгоритмической теории информации, в которой количество информации вводится, исходя из комбинаторных соображений.

Шенноновская теория информации базируется на теории вероятностей. По утверждению А.Н. Колмогорова, реальная сущность энтропии основывается на чисто комбинаторных предположениях, которые несравненно более слабые, чем привлеченные К. Шенноном вероятностные предположения. Основная идея А.Н. Колмогорова заключалась в том, что теория информации должна предшествовать теории вероятностей, а не основываться на ней, поскольку в отличие от последней основания теории информации по самой своей сути должны иметь конечный комбинаторный характер.

Одним из способов такого определения количества информации является ее определение на основе формального понятия алгоритма, в качестве которого чаще всего привлекается машина Тьюринга.

Тогда количество информации, содержащейся в некотором наборе данных, понимается как длина минимальной программы, которая способна породить этот набор данных.

Выбор модели источника информации в рамках данного подхода осуществляется путем обращения идеи оптимального кодирования: в шенноновской теории из знания модели источника сообщений выводились оптимальные коды, здесь же наилучшая модель определяется как модель, обеспечивающая оптимальное кодирование.

Из этих соображений может быть выведен принцип минимальной длины описания: наилучшей моделью является та модель, которая позволяет минимизировать сумму:

- длины описания модели;
- длины описания данных в рамках модели.

Существует несколько близких теоретико-информационных

подходов, опирающихся на алгоритмическую теорию информации, однако в методах анализа изображений длины описаний вычисляются без прямого обращения к данной теории [45]. Вместо этого используются явные представления изображений, задающие алгоритмы восстановления исходных изображений по их описаниям. По сути, представление задает априорное распределение вероятностей на множестве решений, а длина описания изображения заменяет значение правдоподобия [24].

Во многих случаях использование данного подхода оказывается эффективнее, чем явное задание априорных вероятностей $P(h)$ и формы распределения $P(f|h)$. Привлечение понятия представления изображений позволяет объединять этот подход естественным образом с низкоуровневыми, признаковыми, контурными и структурными методами. В области анализа изображений принцип МДО стал активно применяться, начиная с 1990-х годов, и область его применения расширяется. Перечислим ряд примеров:

- Сегментация (по текстуре или цвету) изображения [62, 84].
- Выделение признаков на изображении [45].
- Построение структурных элементов изображений и их группирование, а также описание формы границ областей.
- Распознавание объектов на изображении и распознавание рукописных символов [29].
- Оценивание параметров пространственного преобразования между парой изображений одной и той же сцены, снятой с разных ракурсов, по набору опорных точек и собственно сопоставление и совмещение пары изображений (нахождение соответствия между точками изображений).
- Оценивание движения объекта по видеосерии.

Перспективным признается использование принципа минимальной длины описания (МДО) в задаче выявления изменений, а также в ряде других задач анализа изображений. В то же время на настоящий момент методология использования принципа МДО является заметно менее развитой, чем методология байесовского подхода.

Широкое разнообразие методов анализа изображений объективно обусловлено большим количеством предметных областей, в которых эти методы применяются. Изображения в разных предметных областях могут варьироваться как по своему содержанию, для описания которого может оказаться эффективным то или иное представление изображений, так и по степени изменчивости изображений (типичными причинами изменчивости

изображений являются смена ракурса съемки, освещения, типа камеры, а также собственная изменчивость объектов).

В узких предметных областях с низкой изменчивостью изображений наиболее практичными могут оказаться признаковые методы. Глобальные признаки позволяют осуществлять быструю категоризацию изображений или распознавание изолированных объектов. Простыми в реализации являются корреляционные методы, которые, в отличие от признаковых методов, позволяют распознавать объект на изображении, содержащем несколько объектов, однако корреляционные методы более чувствительны к геометрическим и яркостным преобразованиям изображений. Методы обеих указанных групп обладают низкой степенью инвариантности к разнообразным факторам изменчивости изображений, в связи с чем являются недостаточно универсальными.

Структурные методы обладают наибольшей степенью инвариантности и применимы к широким предметным областям с большой изменчивостью изображений. В то же время при построении структурных описаний происходит, как правило, потеря большого объема информации, в связи с чем эти методы не применимы к малоразмерным объектам. Структурное распознавание изображений можно также оказаться весьма ресурсоемким из-за необходимости перебора большого числа вариантов сопоставления структурных элементов. Однако при наличии перспективных искажений или сложных типов изменчивости объектов на изображениях альтернативу структурным методам найти затруднительно. Повышение эффективности структурных методов может быть достигнуто за счет дополнительного использования локальных текстурных и цветовых признаков. Компромисс между низкоуровневыми и структурными методами представляют контурные методы распознавания изображений, которые допускают достаточно простую и эффективную реализацию. Наиболее существенным недостатком контурных методов является трудность их применения в случаях сложных пространственных преобразований изображений.

Перспективными для построения универсальных систем распознавания изображений могут оказаться иерархические методы [24], включающие несколько взаимодействующих уровней описания изображений. Однако их применение ограничено слабой проработанностью теории создания подобных иерархических систем.

Критерий выбора решения также является важным компонентом методов распознавания. С возрастанием сложности используемого представления повышается сложность задания адекватного критерия, и возрастает влияние критерия на эффективность метода распознавания. В простейших случаях могут

быть использованы эвристические критерии, введение которых требует наименьших трудозатрат. Байесовский подход является наиболее разработанным и может быть рекомендован к использованию на практике в большинстве случаев. Однако для наиболее сложных случаев (как по используемым представлениям, так и по степени изменчивости изображений и широты предметной области) перспективным представляется задание критерия выбора решения на основе принципа минимальной длины описания.

Существует множество задач, требующих применения методов распознавания изображений в информационно-телекоммуникационных системах. Можно выделить два класса задач в зависимости от того, являются ли распознаваемые объекты произвольными или относятся к четко определенному типу (лицам, зданиям, иероглифам и т.д.). Задачи первого класса, возникающие, например, при проведении видеоконференций или при поиске антропогенных объектов в гео-информационных системах, имеют решения приемлемого качества. Для получения этих решений привлекаются корреляционные, признаковые, контурные или структурные методы в зависимости от типа распознаваемых объектов. К примеру, признаковые методы часто применяются в задачах распознавания лиц, поиска местности заданного типа на аэрокосмических изображениях, категоризации изображений и т.д. Контурные и структурные методы чаще используются для распознавания иероглифов, поиска объектов искусственного происхождения на аэрокосмических снимках, поиска изображения по цифровой карте местности в гео-информационных системах или объекта по чертежу в конструкторских базах данных, поиска изображения по наброску, выполненному от руки.

Задачи второго класса связаны с построением универсальных систем распознавания изображений. Такие задачи возникают при построении систем информационного поиска в домашних фотоальбомах, базах торговых знаков, фотостоках, сети Интернет, где изображения могут обладать произвольным содержанием. На настоящий момент большинство функционирующих систем подобного типа используют текстовые аннотации или ключевые слова, на основе которых производится поиск. К примеру, изображения на фотостоках обычно содержат порядка 50 ключевых слов, по совпадению с которыми пользователи, намеривающиеся приобрести фотографии, осуществляют поиск. Тот же принцип используется при проверке уникальности торговых знаков. В результате, пользователи оказываются перед необходимостью просмотра сотен и тысяч изображений в поисках требуемого изображения, которого может и не оказаться в отобранном по ключевым словам наборе. Кроме того, необходимость составления

текстовых аннотаций и списков ключевых слов осложняет процесс создания соответствующих баз изображений.

К сожалению, существующие методы распознавания изображений (среди которых привлекаются признаковые, контурные и структурные методы) лишь в незначительной степени позволяют облегчить данную проблему. Тем не менее, применение методов анализа изображений к задачам, не включающим ограничения на типы распознаваемых изображений, являются наиболее перспективными, но требующим развития как теории построения иерархических представлений изображений, так и теории задания критерия выбора решения в методах распознавания, использующих подобные представления.

2.9. Задача распознавания лиц

При всём многообразии различных алгоритмов распознавания изображений, в основе метода распознавания лежат следующие три основных этапа:

1. преобразование исходного изображения в начальное представление (может включать в себя как предобработку, так и математические преобразования, например вычисление главных компонент);
2. выделение ключевых характеристик (например первые n главных компонент или коэффициентов дискретного косинусного преобразования);
3. механизм классификации (моделирования): кластерная модель, метрика, нейронная сеть и т.п.

Кроме этого, построение метода распознавания опирается на априорную информацию о предметной области (в данном случае – характеристики лица человека), и корректируется экспериментальной информацией, появляющейся по ходу разработки метода.

2.9.1. Метод главных компонент

Метод главных компонент (Principal Component Analysis, **PCA**) применяется для сжатия информации без существенных потерь информативности [11]. Он состоит в линейном ортогональном преобразовании входного вектора X размерности N в выходной вектор Y размерности M , $N < M$. При этом компоненты вектора Y являются некоррелированными и общая дисперсия после преобразования остаётся неизменной.

Матрица X состоит из всех примеров изображений обучающего набора. Решив уравнение

$$\Lambda = \Phi^T \Sigma \Phi,$$

получаем матрицу собственных векторов Φ , где Σ – ковариационная матрица для X , а Λ – диагональная матрица собственных чисел. Выбрав из Φ подматрицу Φ_M , соответствующую M наибольшим собственным числам, получим, что преобразование $y = \Phi_M^T \tilde{x}$, где $\tilde{x} = x - \bar{x}$ – нормализованный вектор с нулевым математическим ожиданием, характеризует большую часть общей дисперсии и отражает наиболее существенные изменения X .

Выбор первых M главных компонент разбивает векторное пространство на главное (собственное) пространство

$$F = \{\Phi_i\}_{i=1}^M,$$

содержащее главные компоненты, и его ортогональное дополнение

$$\bar{F} = \{\Phi_i\}_{i=M+1}^N$$

Применение такого подхода для задачи распознавания изображений лиц заключается в следующем.

Входные вектора представляют собой отцентрированные и приведенные к единому масштабу изображения лиц. Собственные вектора, вычисленные для всего набора изображений лиц, называются собственными лицами (*eigenfaces*). Метод главных компонент в применении к изображениям лиц так же называют методом собственных лиц.

Собственные лица имеют полезное свойство, заключающееся в том, что изображение, соответствующее каждому такому вектору имеет лицеподобную форму (см. рис. 2.5). С помощью вычисленных ранее матриц входное изображение разлагается на набор линейных коэффициентов, называемых главными компонентами. Сумма главных компонент, умноженных на соответствующие собственные вектора, является реконструкцией изображения (рис. 2.6).



Рис. 2.5. Пример изображений собственных векторов (собственные лица)

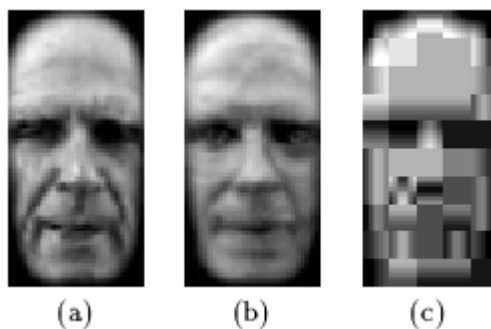


Рис. 2.б.. а) выровненное изображение лица, б) реконструкция по 85-и главным компонентам, в) JPEG - реконструкция (530 байт)

Для каждого изображения лица вычисляются его главные компоненты. Обычно берётся от 5 до 200 главных компонент. Остальные компоненты кодируют мелкие различия между лицами и шум. Процесс распознавания заключается в сравнении главных компонент неизвестного изображения с компонентами всех остальных изображений. Для этого обычно применяют какую-либо метрику (простейший случай – Евклидово расстояние). При этом предполагается что изображения лиц, соответствующих одному человеку, сгруппированы в кластеры в собственном пространстве. Из базы данных (или тренировочного набора) выбираются изображения-кандидаты, имеющие наименьшее расстояние от входного (неизвестного) изображения.

Дальнейшее совершенствование заключалось в использовании метрики Махаланобиса и Гауссовского распределения для оценки близости изображений [11]. Для учёта различных ракурсов в этой же работе использовалось многомодальное распределение изображений в собственном пространстве. Дополнительное повышение надёжности достигалось за счёт дополнительного применения анализа главных компонент к отдельным участкам лица, таким как глаза, нос, рот. Так же метод главных компонент применяется для обнаружения лица на изображении [91]. Для лиц значения компонент в собственном пространстве имеют большие значения, а в дополнении собственного пространства – близки к нулю. По этому факту можно обнаружить, является ли входное изображение лицом. Для этого проверяется величина ошибки реконструкции, чем больше ошибка, тем больше вероятности, что это не лицо.

Исследователями отмечен факт, что при наличии в наборе изображений лиц вариаций, таких как раса, пол, эмоции, освещение, будут появляться компоненты, величина которых в основном определяется этими факторами. Поэтому по значениям соответствующих главных компонент можно определить, например, расу или пол человека.

При изменении ракурса изображения, наступает момент, когда этот метод при распознавании начинает реагировать больше на

ракурс изображения, чем на межклассовые отличия. Классы при этом больше не являются кластерами в собственном пространстве. Это решается добавлением в обучающую выборку изображений в различных ракурсах. При этом собственные вектора теряют лицеподобную форму. В работе [65], развивающей эту идею, показано, что при изменении угла поворота головы, главные компоненты вычерчивают кривые в собственном пространстве, которые однозначно идентифицируют лицо человека и по которым можно провести распознавание. Эти кривые были названы собственными сигнатурами (*eigensignatures*). Отмечалось, что в сочетании с методами генерации изображений в новых ракурсах по одному примеру изображения [93], этот метод имеет неплохие перспективы. По максимумам собственных сигнатур было так же отмечено, что наибольшую информативность имеет изображение лица в полупрофиль.

Аналогичные трудности имеют место при изменении условий освещения. Одна из попыток решения этой проблемы описана в следующем параграфе.

Вычисление набора собственных векторов отличается высокой трудоёмкостью. Один из способов – это свёртка изображений по строкам и столбцам, и дальнейшая работа с полученными результатами [78]. В такой форме представление изображения имеет на порядок меньший размер, вычисления и распознавание происходит быстрее, но восстановить исходное изображение уже невозможно.

Основное преимущество применения анализа главных компонент – это хранение и поиск изображений в больших базах данных, реконструкция изображений [11].

Основной недостаток – высокие требования к условиям съёмки изображений. Изображения должны быть получены в близких условиях освещённости, одинаковом ракурсе и должна быть проведена качественная предварительная обработка, приводящая изображения к стандартным условиям (масштаб, поворот, центрирование, выравнивание яркости, отсечение фона). Нежелательно наличие таких факторов, как очки, изменения в причёске, выражении лица и прочих внутриклассовых вариаций.

Метод собственных лиц требует для своего применения идеализированных условий, таких как единые параметры освещённости, нейтральное выражение лица, отсутствие помех вроде очков и бород. Эти условия в общем случае нельзя достичь путём предварительной обработки. При несоблюдении этих условий главные компоненты не будут отражать межклассовые вариации, и классы перестают представлять собой <кластеры в собственном пространстве. Например, при различных условиях освещённости, метод собственных лиц практически неприменим, поскольку первые

главные компоненты преимущественно отражают изменения освещения, и сравнение выдаёт изображения, имеющие похожий уровень освещённости.

2.9.2. Линейный дискриминантный анализ

Линейный дискриминантный анализ (линейный дискриминант Фишера [53, 70], Linear Discriminant Analysis, LDA), который описывается ниже, выбирают проекцию пространства изображений на пространство признаков таким образом, чтобы минимизировать внутриклассовое и максимизировать межклассовое расстояние в пространстве признаков. В этих методах предполагается что классы линейно разделимы.

Проецирование на собственное пространство смешивает классы, что делает распознавание невозможным, а линейный дискриминант выбирает проекцию на пространство признаков таким образом, чтобы разделить разные классы.

Матрица W для проецирования пространства изображения на пространство признаков выбирается из следующего условия:

$$W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|}$$

где S_B – матрица межклассовой дисперсии, S_W – матрица внутриклассовой дисперсии.

Может существовать до $c-1$ векторов составляющих базис пространства признаков, где c – общее число классов. С помощью этих векторов пространство изображений переводится в пространство признаков.

Поскольку преобразования непосредственно с матрицей $S_W \in \mathbb{R}^{n \times n}$ затруднительна из-за её размерности, в [53] использовано предварительное уменьшение размерности с помощью метода главных компонент, и затем вычисления производятся в пространстве меньшей размерности:

$$W_{pca} = \arg \max_W \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|}$$

где W_{pca} – матрица для проецирования в пространство меньшей размерности (пространство главных компонент). В указанной работе такой метод был назван лицами Фишера (*Fisherfaces*). Так же как и собственные вектора, изображения базисных дискриминантных векторов имеют лицеподобную форму.

В работе [53] тренировочный набор содержал лица при нескольких базовых условиях освещённости, на основе которых при

помощи линейных комбинаций можно получить любые другие условия освещённости. Отмечена высокая точность распознавания (около 96%) для широкого диапазона условий освещённости, различных выражений лица и наличия или отсутствия очков. Была отмечена низкая распознающая способность метода собственных лиц при аналогичных условиях. Причём применение метода собственных лиц, в котором главные компоненты отвечающие за освещённость не учитывались, всё равно давало намного худший результат, чем дискриминант Фишера.

Однако остаются невыясненными вопросы, применим ли этот метод для поиска в больших базах данных, может ли метод работать, когда в тренировочной выборке для некоторых лиц имеется изображение только в одних условиях освещённости.

Так же в данной работе не производилось изменения ракурса, а эксперименты с изменением освещения проводились без изменения других факторов. Будет ли данный метод работоспособен при таких сочетаниях тоже неизвестно. Как и в методе собственных лиц, здесь тоже нужна качественная предварительная обработка, приводящая изображения к стандартным условиям.

Вышеописанный метод основывается на предположении о линейной разделимости классов в пространстве изображений. В общем случае такое предположение несправедливо.

Нейросетевые методы, предлагающие инструмент для построения сложных разделяющих поверхностей, будут описаны далее.

2.9.3. Синтез объектов линейных классов

Данный метод [93] позволяет синтезировать новые изображения объекта (и в частности, изображения лица) для разных ракурсов. Имеется тренировочный набор изображений лиц и только одно изображение нового объекта в определённом ракурсе. Тренировочный набор состоит из изображений объектов того же класса (класс лиц в данном случае), что и новый объект и включает в себя изображения различных лиц, при чём для каждого лица имеются его изображения в широком диапазоне ракурсов. Для нового объекта, имеющего изображение X^A в ракурсе A , осуществляется линейное разложение на изображения объектов из тренировочного набора в том же ракурсе, с вычислением коэффициентов α_i :

$$X^A = \sum_{i=1}^q \alpha_i X_i^A,$$

где q – количество объектов в тренировочном наборе. Синтез

изображения X^B в новом ракурсе B для нового объекта осуществляется сложением изображений из тренировочного набора в ракурсе B с теми же коэффициентами:

$$X^B = \sum_{i=1}^q \alpha_i X_i^B$$

Таким образом метод позволяет синтезировать изображения нового объекта в различных ракурсах по изображению в одном ракурсе без привлечения сложных трёхмерных моделей.

В работе [93] такой метод применялся для синтеза изображений лиц в новых ракурсах. На первом шаге использовалось разложение и синтез изображения с использованием только информации о форме изображения, на втором шаге на изображение в новом ракурсе накладывались текстуры, причём точки соответствия текстур вычислялись с помощью оптического потока.

Данный метод является перспективным для синтеза изображений в новых ракурсах без привлечения сложных трёхмерных моделей, однако вопрос о качестве и количестве примеров в тренировочном наборе остаётся открытым [93].

2.9.4. Гибкие контурные модели лица

В данных методах распознавание производится на основе сравнения контуров лица. Контур обычно извлекается для линий головы, ушей, губ, носа, бровей и глаз. Контур представлен ключевыми позициями, между которыми положение точек, принадлежащих контуру вычисляются интерполированием. Для локализации контуров в различных методах используется как априорная информация, так и информация полученная в результате анализа тренировочного набора.

В работе [81] (*Flexible Models*) ключевые точки размещались вручную на наборе тренировочных изображений. Затем извлекалась информация об интенсивности пикселей, лежащих на линии, перпендикулярной контуру для каждой точки контура. При поиске контуров нового лица использовался метод симуляции отжига с целевой функцией из двух составляющих. Первая из них максимизировалась при соответствии интенсивностей пикселей, извлечённых на перпендикулярной контуру линии аналогичным пикселям из тренировочной выборки. Вторая – при совпадении контура с формой контуров тренировочных примеров. Таким образом, извлекался не просто контур, а контур черт лица. Как должен выглядеть типичный контур черт лица, процедура поиска знала из тренировочных примеров. Для сравнения изображений

использовались значения главных компонент, вычисленных на наборе векторов, представляющих собой координаты ключевых точек. В данной работе контурная модель использовалась вместе с полутоновой моделью, совместное их использование повышало точность распознавания.

Существуют так же другие работы, использующие аналогичные принципы извлечения контуров. Например, в работе [60] использовался генетический алгоритм для получения контуров глаз. Хромосомы представляли собой параметры контура, которые инициализировались таким образом, чтобы начальная область содержала изображение глаза. При этом в функцию оценки пригодности так же была заложена информация о том, как выглядит типичный глаз.

Главной задачей при распознавании по контурам является правильное выделение этих контуров. В общем виде эта задача по сложности сравнима непосредственно с распознаванием изображений. Кроме того, использование этого метода самого по себе для задачи распознавания недостаточно.

2.9.5. Сравнение эластичных графов

В этом методе (*Elastic Bunch Graph Matching*) [95] лицо представляется в виде графа, вершины которого расположены на ключевых точках лица, таких как контуры головы, губ, носы и их крайних точках (рис. 2.7). Каждая грань помечена расстояниями между её вершинами. В каждой такой точке вычисляются коэффициенты Габоровых функций для пяти различных частот и восьми ориентаций. Набор таких коэффициентов $\{J_j\}$ называется **джетом** (*jet*). Джеты характеризуют локальные области изображений и служат для двух целей. Во первых, для нахождения точек соответствия в заданной области на двух различных изображениях. Во вторых – для сравнения двух соответствующих областей различных изображений.

Каждый коэффициент

$$J_j = a_j \exp(i \phi_j)$$

для точек из одной области различных изображений, характеризуется амплитудой a_j , которая медленно меняется с изменением положения точки и фазой ϕ_j , которая вращается со скоростью, пропорциональной частоте волнового вектора базисного вейвлета. Поэтому в простейшем случае для поиска на новом изображении

точки с аналогичными характеристиками в функции подобия фазу не учитывают:

$$S_2(J, J') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a'_j^2}}$$

Функция подобия с одним джетом в фиксированной позиции и другим с переменной позицией является достаточно гладкой, для того чтобы получить быструю и надёжную сходимость при поиске с применением простейших методов, таких как диффузия или градиентный спуск [95]. Более совершенные функции подобия привлекают информацию о фазе.

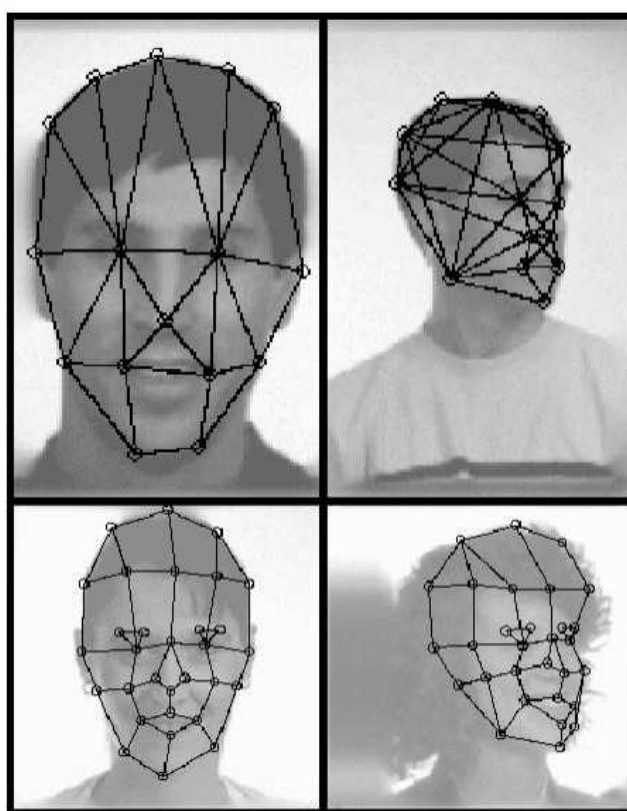


Рис. 2.7. Эластичный граф, покрывающий изображение лица

Для различных ракурсов соответствующие ключевые точки отмечены вручную на тренировочном наборе. Кроме того, чтобы для одного и того же лица представить различные вариации его изображения в одном и том же графе, для каждой точки используются несколько джетов, каждый из которых может соответствовать различным локальным характеристикам данной точки, например открытому и закрытому глазу.

Процесс распознавания неизвестного лица состоит в сравнении графа изображения лица G^I со всеми остальными графами из набора \mathcal{B} при помощи функции подобия:

$$S_B(G^I, B) = \frac{1}{N} \sum_n \max_m (S_\psi(J_n^I, J_n^{Bm})) - \frac{\lambda}{E} \sum_e \frac{(\Delta \bar{x}_e^I - \Delta \bar{x}_e^B)^2}{(\Delta \bar{x}_e^B)^2}$$

Левая сумма характеризует подобие джетов вычисленное с применением фазочувствительной функции, правая – топографическое соответствие, которое пропорционально квадрату разности расстояний между соответствующими вершинами сравниваемых изображений, N – количество вершин, E – количество граней, λ – коэффициент относительной важности топографической информации.

В представленном выше виде метод способен достаточно надёжно распознавать при изменениях ракурса до 22° ; при больших углах точность распознавания резко уменьшается, функция подобия оказывается больше чувствительной к ракурсу, чем к межклассовым различиям. Изменения условий освещённости в работе [95] не производились.

Дальнейшее развитие метода заключается в извлечении коэффициентов важности на основе анализа обучающей выборки [ElWeights]. Для каждого джета симплекс-методом вычисляется коэффициент важности, который затем используется в функции подобия. Коэффициенты важности вычисляются из условия максимизации функции подобия для одного и того же лица и минимизации – для различных лиц.

Однако даже с привлечением коэффициентов важности, вышеописанный метод учитывает только суммарную взвешенную меру геометрических искажений, никак не пытаясь построить модель или учесть характер этих изменений. Другими словами, это перспективный метод извлечения и поиска характеристик на изображении, но этому методу не хватает механизма классификации. Другое улучшение метода заключалось в применении линейных преобразований джетов для компенсации изменения ракурса [76].

Существуют так же более ранние разновидности этого метода, которые не используют изначально определённые ключевые точки и структуры графа [63,66, 96]. Одни из них используют для сравнения решётки джетов, наложенные на изображение (рис. 2.8). В неизвестном изображении отыскиваются точки соответствия, и затем по найденным точкам строится искажённая решётка и измеряется мера её искажения для определения наиболее похожего изображения. В других методах, точки извлечения джетов изначально образуют решётку, а затем наименее пригодные для распознавания точки отсеиваются в процессе обучения.

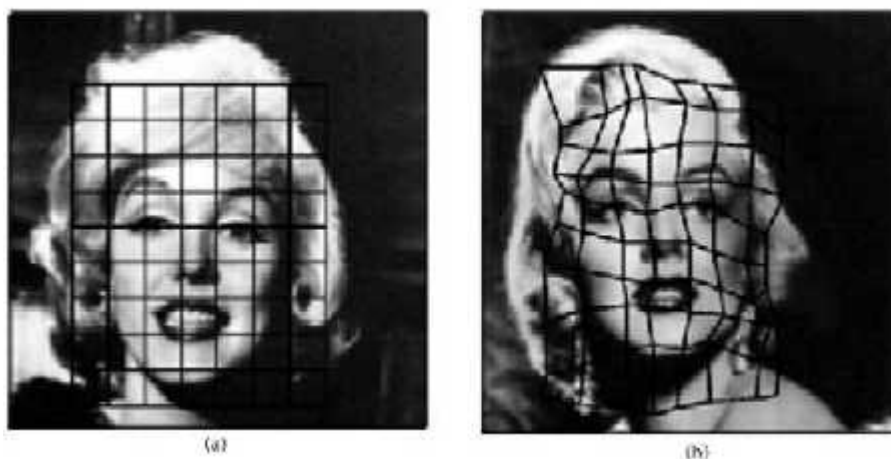


Рис. 2.8. Эластичная решётка, наложенная на изображение, и её искажение

2.9.6. Методы, основанные на геометрических характеристиках лица

Один из самых первых методов – это анализ геометрических характеристик лица [31]. Изначально применялся в криминалистике и был там детально разработан. Потом появились компьютерные реализации этого метода. Суть его заключается в выделении набора ключевых точек (или областей) лица и последующем выделении набора признаков. Каждый признак является либо расстоянием между ключевыми точками, либо отношением таких расстояний. В отличие от метода сравнения эластичных графов, здесь расстояния выбираются не как дуги графов. Наборы наиболее информативных признаков выделяются экспериментально [35].

Ключевыми точками могут быть уголки глаз, губ, кончик носа, центр глаза и т.п. [31], рис. 2.9. В качестве ключевых областей могут быть прямоугольные области, включающие в себя: глаза, нос, рот [67].

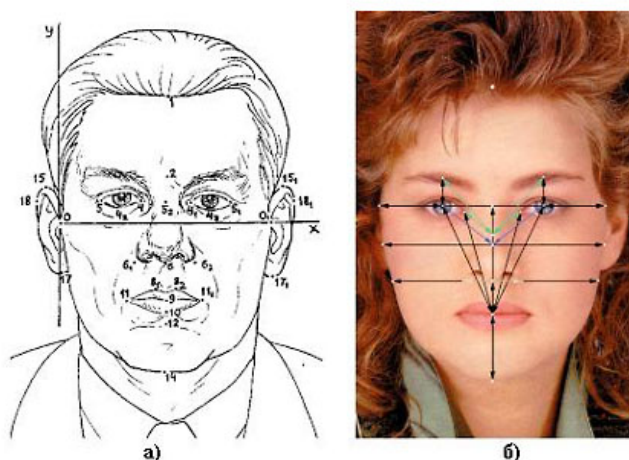


Рис. 2.9. Идентификационные точки и расстояния: а) используемые при криминалистической фотоэкспертизе; б) наиболее часто применяемые при построении автоматизированных систем идентификации.

В процессе распознавания сравниваются признаки неизвестного лица, с признаками, хранящимися в базе.

Задача нахождения ключевых точек приближается к трудоёмкости непосредственно распознавания, и правильное нахождение ключевых точек на изображении во многом определяет успех распознавания [31].

Поэтому изображение лица человека должно быть без помех, мешающих процессу поиска ключевых точек. К таким помехам относят очки, бороды, украшения, элементы причёски и макияжа. Освещение желательно равномерное и одинаковое для всех изображений.

Кроме того, изображение лица должно иметь фронтальный ракурс, возможно с небольшими отклонениями. Выражение лица должно быть нейтральным. Это связано с тем, что в большинстве методов нет модели учёта таких изменений.

Таким образом, данный метод предъявляет строгие требования к условиям съёмки, нуждается в надёжном механизме нахождения ключевых точек для общего случая. Кроме того, требуется применение более совершенных методов классификации или построения модели изменений.

В общем случае этот метод не самый оптимальный, однако для некоторых специфических задач перспективен. К таким задачам можно отнести документный контроль [31, 35], когда требуется сравнить изображение лица, полученного в текущий момент с фотографией в документе. При этом других изображений этого человека не имеется, и следовательно механизмы классификации, основанные на анализе тренировочного набора, недоступны.

2.9.7. Сравнение эталонов

Сравнение эталонов (*Template Matching*) [67] заключается в выделении областей лица на изображении (рис. 2.10), и последующем сравнении этих областей для двух различных изображений. Каждая совпавшая область увеличивает меру сходства изображений. Это так же один из исторически первых методов распознавания человека по изображению лица. Для сравнения областей используются простейшие алгоритмы, вроде попиксельного сравнения.



Рис. 2.10. Сравнимые области-эталоны лица

Недостаток этого метода заключается в том, что он требует много ресурсов как для хранения участков, так и для их сравнения. В виду того, что используется простейший алгоритм сравнения, изображения должны быть сняты в строго установленных условиях: не допускается заметных изменений ракурса, освещения, эмоционального выражения и пр.

2.9.8. Оптический поток

Алгоритмы оптического потока используются в основном для анализа движения. Используя два или более последовательных кадра изображения, можно рассчитать двумерное векторное поле, называемое **оптическим потоком** (*Optical Flow*), которое отражает актуальное или наиболее вероятное смещение точек изображения от кадра к кадру [46, 77].

В работе [77] оптический поток рассчитывался для двух произвольных изображений лица, для получения меры соответствия изображений. Эти два изображения считались последовательными кадрами. Затем вычислялось векторное поле, наилучшим образом отображающее одно изображение в другое, в смысле минимизации расстояния между изображениями и с учётом геометрических ограничений, таких как относительное расположение точек изображения. Алгоритм находил наиболее соответствующие блоки. Поиск осуществлялся иерархически, начиная с больших блоков, и затем разбивая их на меньшие блоки. Таким образом, строилась пирамида соответствия изображений.

Используя векторное поле оптического потока, строились различные меры соответствия, с помощью которых в базе находилось изображение, ближайшее к неизвестному.

База данных представляла собой 76 изображений 36-и человек с различиями в ракурсе и выражении лица. Распознавание осуществлялось несколькими способами.

В первом способе блоки 8x8 неизвестного изображения

заменялись на наиболее соответствующие блоки сравниваемого изображения (рис. 2.11, 2.12). Затем вычислялось Евклидово расстояние между неизвестным и полученным изображением. Было достигнуто 92% точности распознавания. Учитывая то, что в базе находилось только одно изображение нужного человека и по два на всех остальных, это хороший результат.



Рис. 2.11. Отображение неизвестного изображения на известное, один и тот же человек. Слева направо: неизвестное изображение, изображение из базы данных, неизвестное изображение, в котором блоки заменены блоками из известного изображения.



Рис. 2.12. Отображение неизвестного изображения на изображение из базы. Изображения разных людей.

Как оказалось, даже при отображении изображения одного человека на изображение другого, результирующее изображение оказывается визуально очень похожим на исходное (рис. 2.12). Поэтому в качестве другой меры соответствия использовалась дивергенция оптического потока. Точность оказалась хуже, вследствие больших искажений у блоков, не представляющих важности (таких как волосы например). Использование центральной части лица дало немного лучший результат.

К недостаткам этого метода в первую очередь относится его вычислительная трудоёмкость. На специализированном параллельном компьютере сравнение двух изображений занимало больше минуты. Метод неспособен извлекать компактный набор характеристик для хранения и поиска в базе.

Однако само направление представляет большой интерес, надо экспериментировать с более репрезентативными базами изображений лиц.

Метод анализирует только суммарное искажение между изображениями или только суммарное несоответствие блоков, не касаясь характера этих искажений. Метод, анализирующий как характер искажения изображений, так и соответствие отдельных блоков, может дать отличный результат. Это подтверждает работа [58], применяющая псевдодвумерные скрытые Марковские модели.

2.9.9. Скрытые Марковские модели

Марковские модели являются мощным средством моделирования различных процессов и распознавания образов. По своей природе Марковские модели позволяют учитывать непосредственно пространственно-временные характеристики сигналов, и поэтому получили широкое применение в распознавании речи, а в последнее время – изображений (в частности изображений лиц).

Каждая модель $\lambda = (A, B, \pi)$, представляет собой набор N состояний $S = \{S_1, S_2, \dots, S_N\}$ (вершины графа), между которыми возможны переходы (дуги) [88]. В каждый момент времени система находится в строго определённом состоянии. В наиболее распространённых Марковских моделях **первого порядка** полагается, что следующее состояние зависит только от текущего состояния.

При переходе в каждое состояние генерируется наблюдаемый символ, который соответствует физическому сигналу с выхода моделируемой системы. Набор символов для каждого состояния $V = \{v_1, v_2, \dots, v_M\}$, количество символов M . Выход, генерируемый моделью, может быть так же непрерывным. Существуют так же модели, в которых набор символов для всех состояний одинаков.

Символ в состоянии $q_t = S_j$ в момент времени t генерируется с вероятностью $b_j(k) = P[v_{k,t} | q_t = S_j]$. Набор всех таких вероятностей составляет матрицу $B = \{b_j(k)\}$.

Матрица $A = \{a_{ij}\}$ определяет вероятность перехода из одного состояния в другое состояние: $a_{ij} = P[q_{t+1} = S_j | q_t = S_i]$, $1 \leq i, j \leq N$. Считается, что A не зависит от времени. Если из каждого состояния можно достичь любого другого за один переход, то все $a_{ij} > 0$, и модель называется эргодической. Так же модель имеет вероятность начальных состояний $\pi = \pi_i$, где $\pi_i = P[q_1 = S_i]$.

Обычно в реальных процессах последовательность состояний

является скрытой от наблюдения и остаётся неизвестной, а известен только выход системы, последовательность наблюдаемых символов $O = O_1 O_2 \dots O_T$, где каждое наблюдение O_t – символ из V , и T – число наблюдений в последовательности. Поэтому такие модели называют скрытыми Марковскими моделями (СММ, по английски – Hidden Markov Models, НММ).

Модель $\lambda = (A, B, \pi)$ с настроенными параметрами может быть использована для генерирования последовательности наблюдений. Для этого случайно, в соответствии с начальными вероятностями π выбирается начальное состояние, затем на каждом шаге вероятность B используется для генерации наблюдаемого символа, а вероятность A – для выбора следующего состояния. Вероятность P генерирования моделью λ последовательности состояний Q :

$$P(Q | \lambda) = \prod_{t=1}^T b_{q_t}(O_t),$$

где $Q = q_1 q_2 \dots q_T$ – последовательность состояний. Предполагается, что наблюдения статистически независимы.

В распознавании образов скрытые Марковские модели применяются следующим образом. Каждому классу i соответствует своя модель λ_i . Распознаваемый образ (речевой сигнал, изображение и т.д.) представляется в виде последовательности наблюдений O . Затем для каждой модели λ_i вычисляется вероятность того, что эта последовательность могла быть сгенерирована именно этой моделью. Модель λ_j , получившая наибольшую вероятность, считается наиболее подходящей, и образ относят к классу j .

В связи с этим появляются несколько вопросов, называемых тремя основными задачами скрытых Марковских моделей [22].

1. Имея последовательность наблюдений $O = O_1 O_2 \dots O_T$ и настроенную модель $\lambda = (A, B, \pi)$, как оценить вероятность $P(O | \lambda)$ генерации этой моделью данной последовательности наблюдений? Эта задача называется задачей **расознавания**.

2. Имея последовательность наблюдений $O = O_1 O_2 \dots O_T$ и настроенную модель $\lambda = (A, B, \pi)$, как подобрать последовательность состояний $Q = q_1 q_2 \dots q_T$, чтобы она была оптимальной (в соответствии с некоторым критерием, аналитически эта задача неразрешима)? Другими словами это задача **объяснения**. Она нужна для последующей коррекции параметров модели.

3. Каким образом корректировать параметры модели λ , для того чтобы максимизировать $P(O | \lambda)$? Т.е. как сделать так, чтобы

модель больше соответствовала своему классу, одним из образов которого является данная последовательность наблюдений (или несколько различных последовательностей). Это задача **обучения**.

Первая задача имеет точное аналитическое решение, называемое процедурой прямого-обратного прохода. Последующие две задачи не имеют точного аналитического решения. Для решения второй задачи используется алгоритм Витерби, для третьей – алгоритм Баума-Вельча. Оба этих метода являются разновидностями градиентного спуска и решаются оптимизационными методами.

Для того, чтобы сократить вычисления, в распознавании речи используются линейные модели [88] (рис. 2.13). В таких моделях каждое состояние имеет только одно последующее, так же переход возможен обратно в то же состояние. Такие модели учитывают временные характеристики речевого сигнала: определённый порядок следования участков сигнала, их взаимное расположение, возможность локальных растяжений или сжатий. Это позволяет их применять и в распознавании изображений.

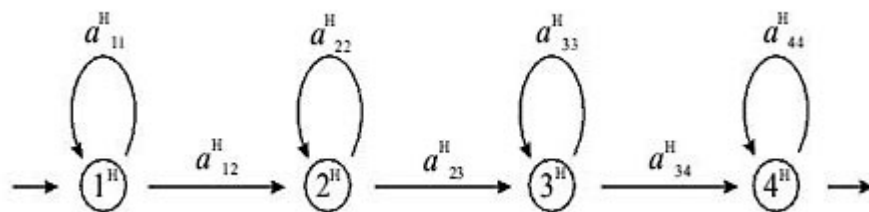


Рис. 2.13. Линейная Марковская модель

Одна из первых работ, применяющая СММ для распознавания изображений лиц [89], которой предшествовали работы по распознаванию изображений других видов скрытыми Марковскими моделями. В этой работе распознавание осуществлялось как простейшими одномерными линейными СММ, так и псевдодвумерными. Введение второго измерения позволило повысить точность распознавания с 85% до 95% на базе ORL.

Суть двумерных Марковских моделей заключается в том, что в отличие от одномерных линейных СММ, они позволяют моделировать искажения изображения и взаимное расположение участков не отдельно по горизонтали или вертикали, а в обоих направлениях одновременно. Для уменьшения вычислительной сложности применяются псевдодвумерные СММ (Pseudo-2D Hidden Markov Models, P2D-HMM). Такая модель состоит из нескольких линейных вертикальных моделей нижнего уровня, и одной линейной горизонтальной модели верхнего уровня, на вход которой поступают выходы моделей нижнего уровня (рис. 2.14). Каждое состояние модели верхнего уровня включает в себя последовательность

состояний соответствующей модели нижнего уровня. Модели нижнего уровня не связаны между собой. Изначально в [89] модели верхнего уровня были вертикальными. В последующих работах [57, 89] модели верхнего уровня были сделаны горизонтальными (как это и изображено на рисунке), в связи с тем, чтобы вертикальные модели нижнего уровня могли учесть факт того, что глаза могут находиться на разной высоте. Таким образом псевдодвумерная модель позволяет учесть локальные деформации и взаимное расположение участков изображений. Но в отличие от оптических потоков и других методов сопоставления деформациями, псевдодвумерная модель учитывает характер деформаций, а то какими именно могут быть возможные деформации, псевдодвумерные СММ усваивают в процессе обучения. Другими словами, участок, соответствующий глазу, никогда не будет сопоставлен например участку на месте рта, как это может быть в оптическом потоке.

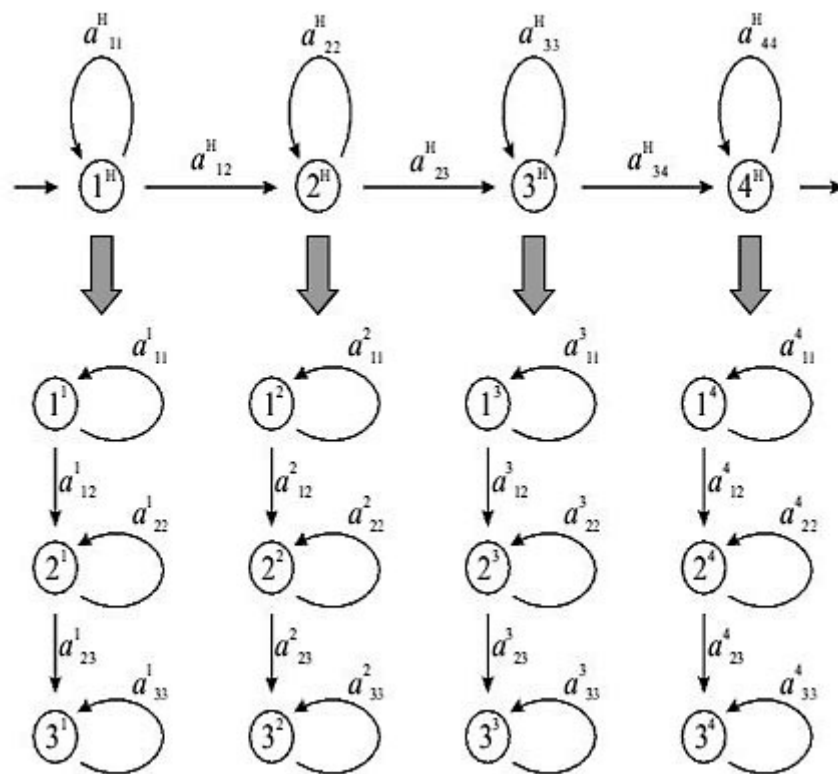


Рис. 2. 14. Псевдодвумерная скрытая Марковская модель

Наблюдениям, подаваемым на вход СММ, являлись квадратные участки изображений (рис. 2.15). Было обнаружено, что участки, извлекаемые с 75% перекрытием друг с другом, давали наилучшую точность распознавания.

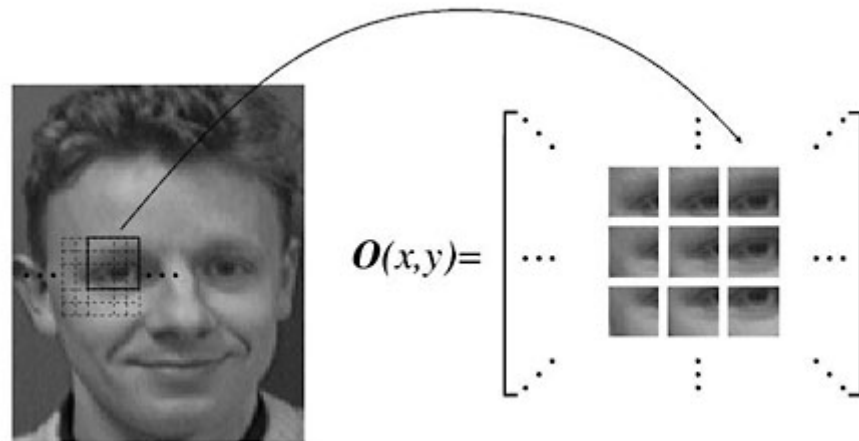


Рис. 2.15. Извлечение участков-образцов наблюдения

Одним из полезных свойств СММ является способность сегментировать распознаваемое изображение. Результат работы алгоритма Витерби, разбившего изображение на последовательность состояний, показан на рис. 2.16. [

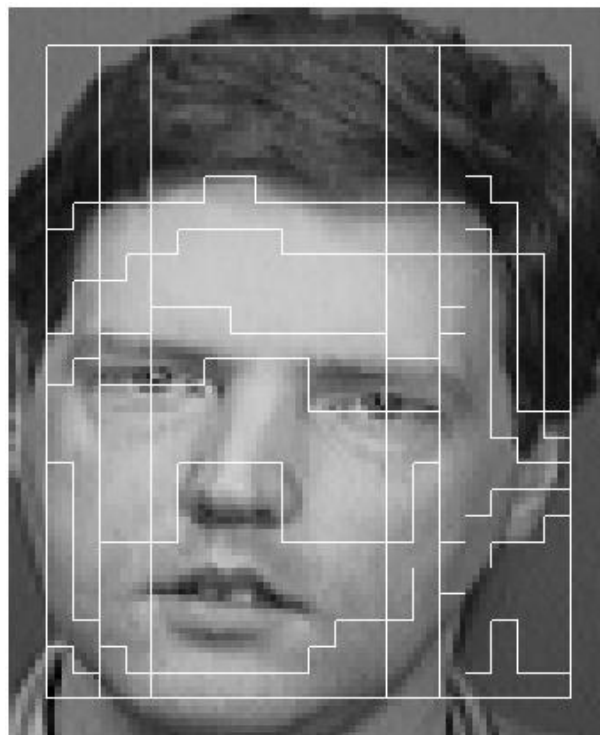


Рис. 2.16. Сегментация изображения. Линии отмечают области, соответствующие одинаковым состояниям.

В работе [57], продолжающей идею Самарии использовались дальнейшие улучшения способов начального представления изображения и алгоритмов тренировки.

Для каждого квадратного участка изображения 16×16 вычислялось двумерное дискретное косинусное преобразование, и этот участок представлялся в виде набора первых 15-ти коэффициентов ($u+v \leq 4$, п. 2.7.2.). Это позволило повысить точность

распознавания на 2%. Кроме того, такое представление позволяет более точно, чем при масштабировании представлять изображение, используя меньший объём информации. Для увеличения тренировочного набора использовались так же зеркально отражённые по вертикали изображения [57]. Это позволило учесть более широкий диапазон ракурсов.

Для СММ важное значение имеет начальная инициализация модели. В [57], в качестве начальной инициализации всех моделей использовались все изображения из тренировочного набора. Затем модель каждого класса донастраивалась на свои изображения. На базе ORL было достигнуто 100% точность распознавания. Авторы этой работы считают это основным фактором, давшим прирост производительности.

Полезное свойство распознавания по коэффициентам дискретного косинусного преобразования заключается в том, что оно позволяет работать непосредственно со сжатыми изображениями, такими как JPEG и MPEG, в которые на сегодняшний день являются распространёнными форматами хранения изображений и видео. В работе [57] производились эксперименты по распознаванию сжатых JPEG <изображений. Было достигнуто 99.5% точности на той же базе. Падение точности связано с невозможностью извлекать перекрывающиеся блоки в процессе распознавания (но в процессе обучения перекрывающиеся блоки были использованы).

Как можно видеть, правильный выбор метода классификации (моделирования), начального представления изображения и учёт его особенностей в сочетании с применением априорных знаний о предметной области, дало успешный результат.

Недостатком СММ является то, что СММ не обладает **различающей** способностью. Т.е. алгоритм обучения только максимизирует отклик каждой модели на свои классы, но не минимизирует отклик на другие классы, и не выделяются ключевые признаки, отличающие один класс от другого. Например, для определения того, содержится ли лицо в обучающей выборке, использовался алгоритм ранжирования вероятностей, заключавшийся в следующем.

На обучающем наборе, каждая модель реагирует на изображения-примеры с некоторой вероятностью. Отсортированные таким образом модели образуют исходное ранжирование. Для неизвестного изображения модели так же ранжируются по вероятностям отклика на неизвестное изображение. Большая величина отклонения полученного ранжирования от исходного сигнализирует о том, что изображение принадлежит к неизвестному классу лица.

Таким образом, похожие классы могут оказаться слабо

различимыми с единственной ошибкой) и при увеличении объёма базы или использования в более широких условиях СММ может оказаться ненадёжными. Многослойные нейронные сети лишены такого недостатка.

3. ПРИКЛАДНЫЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

В настоящее время всё более широкое распространение получают биометрические системы идентификации человека. Традиционные системы идентификации требуют знания пароля, наличия ключа, идентификационной карточки, либо иного идентифицирующего предмета, который можно забыть или потерять. В отличие от них биометрические системы основываются на уникальных биологических характеристиках человека, которые трудно подделать и которые однозначно определяют конкретного человека. К таким характеристикам относятся отпечатки пальцев, форма ладони, узор радужной оболочки, изображение сетчатки глаза. [10,18,21]

События последнего времени, особенно террористический акт в Нью-Йорке, безусловно, существенно повлияли на мировой рынок биометрии. Так например, среди граждан США всего 10% поддерживало идею биометрической паспортизации до 11 сентября 2001 года и уже более 75% — после теракта, когда отслеживание потенциально опасных личностей стало первостепенной задачей. Биометрия получила весомый шанс стать одной из самых быстроразвивающихся отраслей безопасности.

Сегодня мировой рынок биометрических систем формируют более 300 компаний, которые занимаются разработкой, производством, продажей и обслуживанием средств и систем безопасности. Структура этого рынка весьма досконально была изучена специалистами фирмы IDC и по состоянию на 2012 года приведена на рис. 3.1.

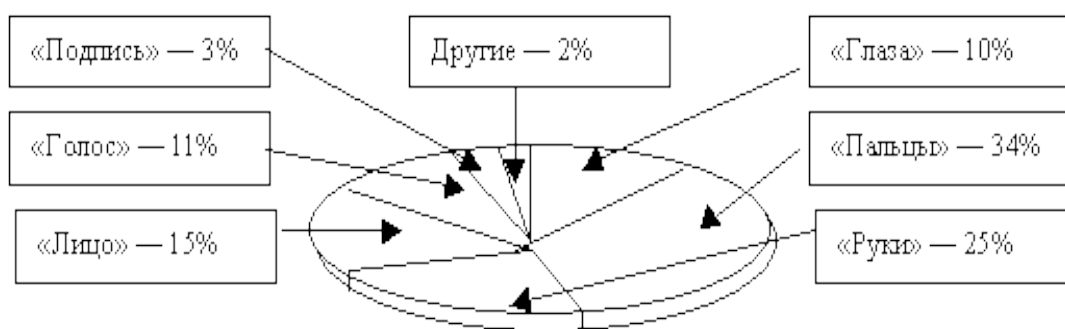


Рис. 3.1. Структура мирового рынка биометрических средств защиты

Анализ рынка четко показывает, что основной интерес покупателя составляют технологии и средства контроля доступа в здания и к компьютерам. Средний годовой темп развития биометрии составляет 40%, что является высоким показателем даже для растущей экономики. При сохранении таких темпов всего через 15 лет население Земли будет обеспечено биометрическими

удостоверениями личности, информация о которых будет храниться в государственных базах данных, объединенных в глобальную международную идентификационную систему. Кстати, это предложение, высказанное представителем США после теракта 11 сентября 2001 г., в настоящее время поддержано Россией.

В настоящее время отечественной промышленностью и рядом зарубежных фирм предлагается достаточно широкий набор различных средств контроля доступа к информации, в результате чего выбор оптимального их сочетания для применения в каждом конкретном случае вырастает в самостоятельную проблему. По своему происхождению на российском рынке в настоящее время представлены как отечественные, так и импортные биометрические средства защиты информации (БСЗИ), хотя существуют и совместно разработанные средства. По конструктивным особенностям можно отметить системы, выполненные в виде моноблока, нескольких блоков и в виде приставок к компьютерам. Возможная классификация биометрических средств защиты информации по биометрическим признакам, принципам действия и технологии реализации приведена на рис. 3.2.

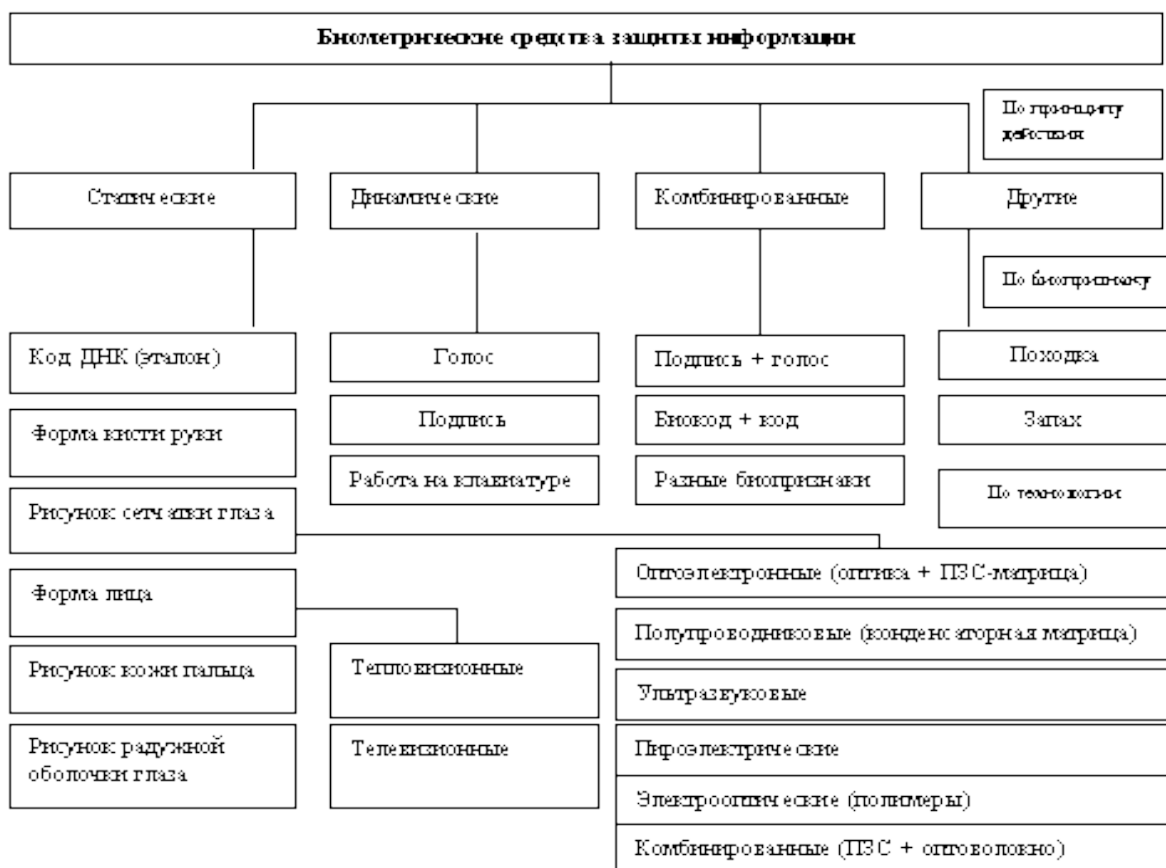


Рис. 3.2. Классификация биометрических средств защиты информации

В настоящее время биометрические системы контроля доступа к информации завоевывают все большую популярность в банках, фирмах, связанных с обеспечением безопасности в телекоммуникационных сетях, в информационных отделах фирм и т. д. Расширение применения систем этого типа можно объяснить как снижением их стоимости, так и повышением требований к уровню безопасности. Подобные системы на российском рынке появились благодаря фирмам “Identix”, “SAC Technologies”, “Eyedentify”, “Biometric Identification Inc.”, “Recognition Systems”, “Trans-Ameritech”, “BioLink”, “Sonda”, “Elsys”, “Эдванс”, “ААМ Системз”, “Полми групп”, “Маском”, “Биометрические системы” и др.

Все биометрические системы характеризуются высоким уровнем безопасности, прежде всего потому, что используемые в них данные не могут быть утеряны пользователем, похищены или скопированы. В силу своего принципа действия многие биометрические системы пока еще отличаются сравнительно малым быстродействием и низкой пропускной способностью. Тем не менее, они представляют собой единственное решение проблемы контроля доступа на особо важных объектах с малочисленным персоналом. Например, биометрическая система может контролировать доступ к информации и хранилищам в банках, ее можно использовать на предприятиях, занятых обработкой ценной информации, для защиты ЭВМ, средств связи и т.д. По оценкам специалистов, более 85% установленных в США средств биометрического контроля доступа предназначались для защиты машинных залов ЭВМ, хранилищ ценной информации, исследовательских центров, военных установок и учреждений.

Распознавание человека по изображению лица выделяется среди биометрических систем тем, что во-первых, не требуется специальное или дорогостоящее оборудование. Для большинства приложений достаточно персонального компьютера и обычной видеокамеры. Во-вторых, не нужен физический контакт с устройствами. Не надо ни к чему прикасаться или специально останавливаться и ждать срабатывания системы. В большинстве случаев достаточно просто пройти мимо или задержаться перед камерой на небольшое время.[18]

Основные приложения биометрической идентификации личности по изображению лица к различным сферам человеческой деятельности приведены в таблице 3.1.:

Основное отличие данных приложений между собой целевые классы, которые являются объектами распознавания. Целевыми классами в задачи распознавания лица могут являться: лицо, лицо с элементами перекрытий, мимика лица, пол, раса, личность человека. Выбор одного из таких целевых классов и определяет специфику

алгоритма распознавания, остальные классы являются второстепенными и скорее играют роль признаков при распознавании целевого класса.

К недостаткам распознавания человека по изображению лица следует отнести то, что сама по себе такая система не обеспечивает 100%-ой надёжности идентификации [17,18]. Там, где требуется высокая надёжность, применяют комбинирование нескольких биометрических методов(мультимодальные биометрические системы).

Табл. 3.1.

Применение биометрической идентификации в человеческой деятельности

Сфера деятельности	Приложения
Информационная безопасность	Безопасность доступа (ОС, базы данных) Сохранность личных данных (медицинских и т.п.) Аутентификация пользователя (торговля, он-лайн банкинг)
Контроль доступа	Аутентификация в системах безопасности Системы связанные с закрытым доступом
Биометрика	Идентификация личности (паспорта, водительские права, удостоверения голосующего и т.п.) Автоматическое подтверждение личности (таможенный контроль)
Правоохранительные органы	Видео наблюдение Идентификация подозреваемых Отслеживание подозреваемых в процессе расследований Реконструкция лиц по свидетельским показаниям
Личная безопасность	Домашние системы видео наблюдения Системы контроля состояния человека (например, система анти-сон для водителей)
Развлечения и отдых	Системы интерактивных видеоигр Приложения фото и видео камер

На данный момент проблеме распознавания человека по изображению лица посвящено множество работ, однако в целом она ещё далека от разрешения. Основные трудности состоят в том, чтобы распознать человека по изображению лица независимо от изменения ракурса и условий освещённости при съёмке, а так же при различных изменениях, связанных с возрастом, причёской и т.д. [10,19]

Распознавание изображений пересекается с распознаванием образов.[21] Такие задачи не имеют точного аналитического решения. При этом требуется выделение ключевых признаков, характеризующих зрительный образ, определение относительной важности признаков путём выбора их весовых коэффициентов и учёт взаимосвязей между признаками. Изначально эти задачи выполнялись человеком-экспертом вручную, путём экспериментов, что занимало много времени и не гарантировало качества. В новых методах выделение ключевых признаков осуществляется путём автоматического анализа обучающей выборки, но тем не менее большинство информации о признаках задаётся вручную. Для автоматического применения таких анализаторов выборка должна быть достаточно большой и охватывать все возможные ситуации.

Нейросетевые методы предлагают иной подход к решению задачи распознавания образов.[11,12,20-22] Архитектура и функционирование нейронных сетей (НС) имеют биологические прообразы. Веса в нейронной сети не вычисляются путём решения аналитических уравнений, а подстраиваются различными локальными методами (например разновидностями градиентного спуска) при обучении. Обучаются нейронные сети на наборе обучающих примеров. В процессе обучения НС происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними. Обученная НС может успешно применять опыт, полученный в процессе обучения, на неизвестные образы за счёт хороших обобщающих способностей.

При всём многообразии различных алгоритмов и методов распознавания изображений они имеют схожую структуру. Типичный метод распознавания состоит из трёх компонент показан на рис. 3.3:

преобразование исходного изображения в начальное представление (может включать в себя как предобработку, так и математические преобразования, например вычисление главных компонент);

выделение ключевых характеристик (например берётся первые n главных компонент или коэффициентов дискретного косинусного преобразования);

механизм классификации (моделирования): кластерная модель, метрика, нейронная сеть и т.п.

Кроме этого, построение метода распознавания опирается на априорную информацию о предметной области (в данном случае – характеристики лица человека), и корректируется экспериментальной информацией, появляющейся по ходу разработки метода.

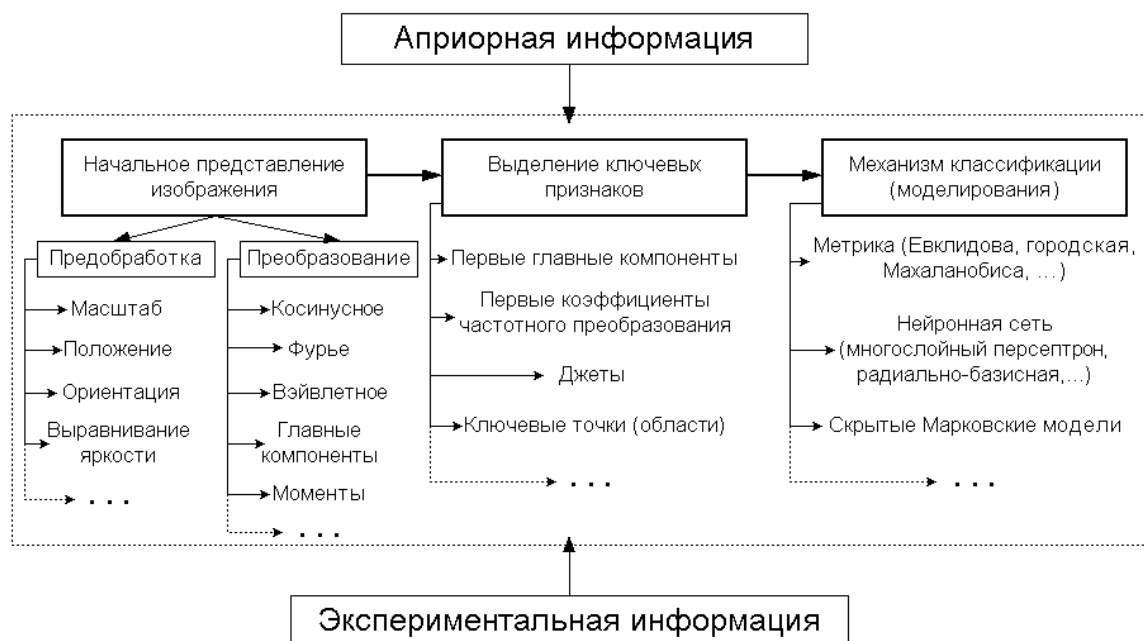


Рис. 3.3. Структура метода распознавания изображений

3.1. Классификация и применение искусственных нейронных сетей.

3.1.1. Основные классы решаемых задач при распознавании человека по изображению лица

Задачи распознавания человека по изображению лица делятся на три больших класса: поиск в больших базах данных, контроль доступа и контроль фотографий в документах. Они различаются как по требованиям, предоставляемым к системам распознавания, так и по способам решения, и поэтому представляют собой отдельные классы.

Различны и требования, предъявляемые к ошибкам первого и второго рода для таких классов [10]. Ошибкой первого рода (type I error, misdetection) называется ситуация, когда объект заданного класса не распознаётся (пропускается) системой. Ошибка второго рода (type II error, false alarm) происходит, когда объект заданного класса принимается за объект другого класса. Следует так же отметить различие понятий верификации и распознавания (идентификации)[17]. В задаче верификации неизвестный объект заявляет, что он принадлежит к некоторому известному системе классу. Система подтверждает или опровергает это заявление. В системах верификации ошибкой первого рода является ситуация, когда объект, принадлежащий к известным системе классам, принимается за объект, относящийся к неизвестным системе классам, и в доступе ему отказывают [18,22]. Ошибка второго рода совершается, когда объект неизвестного класса принимается за

объект, относящийся к известным системе классам, и ему разрешается доступ [18,22]. При распознавании требуется отнести объект к одному из n известных классов или выдать заключение о том, что этот объект не относится к известным классам.

3.1.2. Поиск изображения в больших базах данных

Сравнение типа «один со многими». Высокие требования к ошибке первого рода – система распознавания должна находить изображения, соответствующие данному человеку, по возможности не пропустив ни одного такого изображения. При этом допустимо, если в результирующей выборке будет присутствовать небольшое число других людей.

Обычно в большой базе данных (10^4 - 10^7 изображений) требуется найти изображения, наиболее похожие на заданное [48]. Поиск должен быть произведён за разумное время. Одно из решений состоит в хранении в базе данных небольших наборов заранее извлечённых ключевых признаков, максимально характеризующих изображение. При этом требования к точности не столь критичны, как в задачах контроля доступа и документного контроля. К данному классу прежде всего относится метод главных компонент (метод «собственных лиц») [18,31,32,55]. Коэффициенты, полученные разложением входного изображения на главные компоненты, использовались для сравнения изображений путём вычисления Евклидова расстояния, а в более совершенных методах – на основе метрики Махала-Нобиса с использованием Гауссовского распределения [56].

В работах [52,74] описано развитие метода главных компонент на основе нейронных сетей. В работе [90] также показана возможность использования признаков, сформировавшихся на более поздних слоях специализированной свёрточной нейронной сети, для классификации изображений по методу ближайшего соседа.

На рис. 3.4. показан алгоритм работы системы при поиске информации в базе данных. Система наблюдения делает фотографию человека. С помощью нейронной сети производится поиск области лица на этой фотографии. Область лица выделяется, оптимизируются яркость, контраст изображения, затем нормализованный фотопортрет поступает на обработку второй нейронной сети для распознавания. Нейронная сеть производит распознавание входящего портрета и осуществляет выбор из нескольких наиболее похожих на него портретов, хранящихся в базе данных.

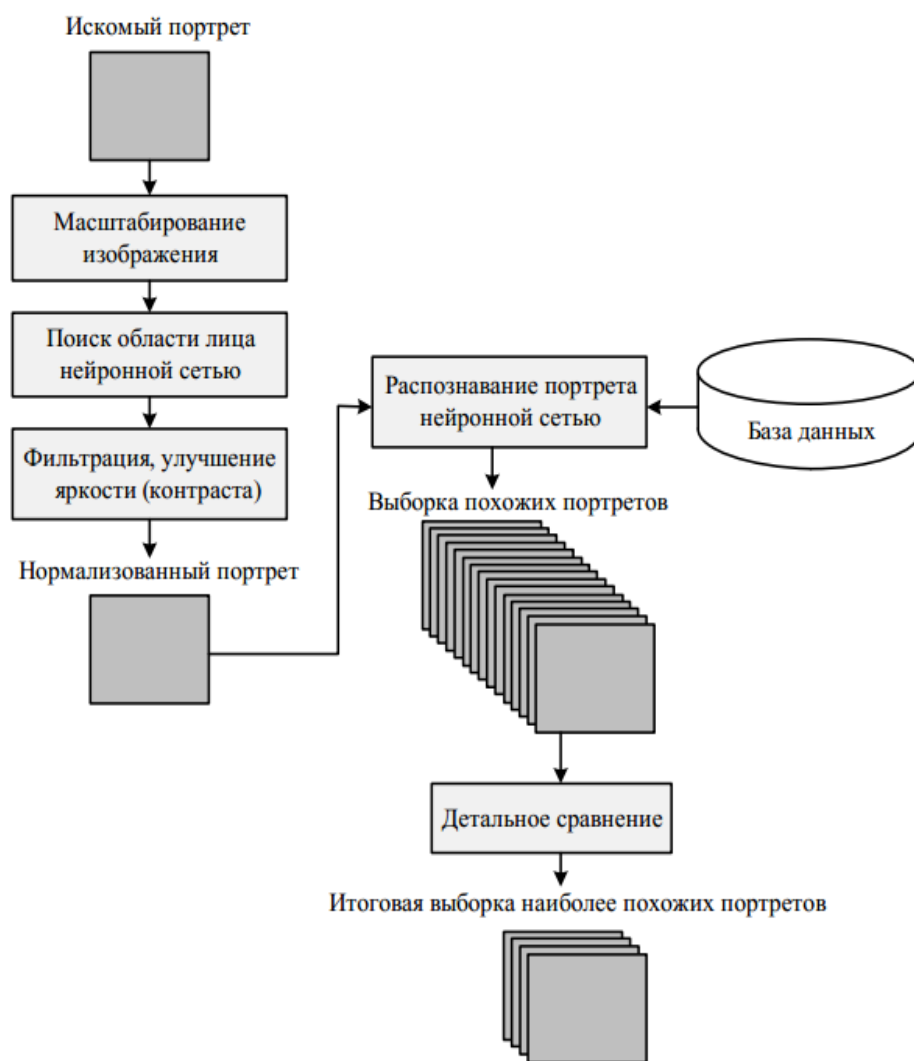


Рис. 3.4. Алгоритм работы системы поиска

3.1.3. Задача контроля доступа

Сравнение типа «один с несколькими». Критическими являются требования к ошибкам второго рода. Система распознавания не должна распознавать незнакомых людей как знакомых, возможно даже за счёт увеличения ошибок первого рода (отказов в доступе знакомым людям).

Имеется небольшая группа лиц (5-50 человек), которых система должна распознавать по изображению лица и открывать им доступ в некоторое помещение. Людей, не входящих в эту группу, система не должна пропускать. Возможны варианты, когда требуется установить конкретную личность по изображению лица. При этом от системы требуется высокая достоверность распознавания, возможно даже за счёт увеличения числа отказов на знакомые объекты. В качестве тренировочных изображений обычно для каждого человека доступны несколько изображений лица, полученных при различных условиях. Это могут быть, например, изменения ракурса, условий

освещённости, причёски, мимики, наличие или отсутствие очков и т.п. Система должна работать в реальном масштабе времени, а процесс настройки может занимать больше времени и выполняться предварительно. В процессе эксплуатации система должна дообучаться на вновь поступающих изображениях по возможности быстрее. Ограничений на применяемые методы здесь нет, но все методы сходятся в том, что имеется обучающий набор изображений лиц заданной группы людей (возможно при различных условиях съёмки). К этому набору система обращается в процессе распознавания или настраивается на него в процессе обучения.

Работу подобной системы контроля доступа можно описать следующим образом. На двери расположена фото- или видеокамера, которая фиксирует человека на входе. Делается фотоснимок человека. На фотоснимке находится область расположения лица, далее происходит распознавание этого лица. Если лицо соответствует портрету, хранящемуся в базе данных, то читается дополнительная информация: имя, возраст, должность и т. д. На основе этих данных система открывает или закрывает доступ к объектам. На рис. 3.5. показан алгоритм работы системы при вводе информации в базу данных. На вход поступает изображение, представляющее собой фотографию человека. С помощью нейронной сети производится поиск области лица на фотографии. Область лица выделяется, оптимизируются яркость, контраст изображения, после чего данный фотопортрет сохраняется в отдельном файле базы данных. Пользователь системы может добавить с клавиатуры дополнительную информацию о человеке, изображенном на фото: возраст, пол, должность и т. д. Эта информация записывается в тот же файл базы данных.

3.1.4. Задача контроля фотографии в документах

Сравнение типа «один с одним». Формулировать требования к ошибкам первого и второго рода как к системе верификации или распознавания здесь будет некорректно, поскольку система распознавания никогда не имела дело с поступающими на вход классами. Но желательно, чтобы система не совершала ошибок при сравнении. Требуется сравнить изображение лица человека, полученное в данный момент, с фотографией из какого-либо документа. Системе надо ответить, принадлежат ли эти лица одному человеку или нет. Данный класс задач наиболее сложен, поскольку, во-первых, система никогда раньше не сталкивалась с изображением лица данного человека. Система сравнивает всегда отличающиеся изображения, учёт всех возможных различий в процессе обучения или настройки системы затруднителен. Во-вторых, здесь большое

влияние оказывают возрастные и другие изменения лица. В-третьих, качество и контраст отсканированной фотографии, как правило, хуже, чем изображение лица, снятого камерой [18,33,47].

Большинство методов для данного класса задач неприменимы без специальной адаптации. В обзоре литературы нет работ, напрямую связанных с применением нейросетевых методов для решения данного класса задач. Для этого можно предложить применение НС для извлечения ключевых признаков изображений и адаптацию НС для сравнения двух изображений.

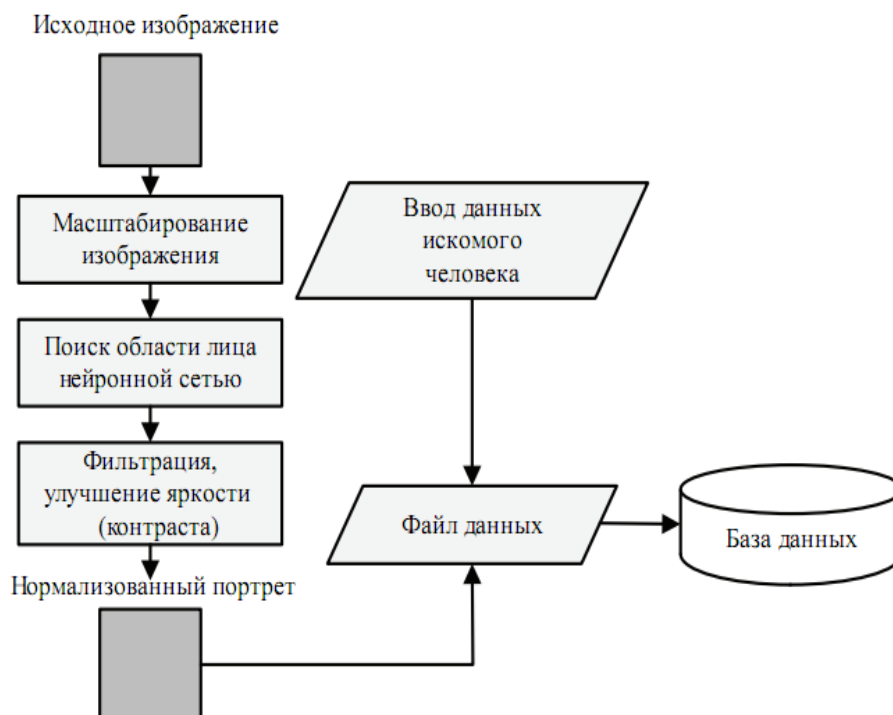


Рис. 3.5. Алгоритм ввода данных в систему контроля доступа.

3.1.5. Нейросетевые методы распознавания человека по изображению лица

Нейросетевые методы, основанные на применении различных типов искусственных нейронных сетей (ИНС, в дальнейшем просто нейронные сети, НС), в последнее время получили широкое распространение. Основные задачи, решаемые при помощи нейронных сетей [20,36,37,49] приведены на рис.3.6.

Большинство из этих задач прямо или косвенно связаны с распознаванием изображений. Основные преимущества, которыми обладают нейронные сети, следующие:

Настройка нейронной сети для решения определённой задачи производится в процессе обучения на наборе тренировочных примеров. Таким образом, не требуется вручную определять параметры модели (выбирать ключевые признаки, учитывать их

взаимоотношение и т.п.) – НС извлекает параметры модели автоматически наилучшим образом в процессе обучения. Остаётся только построить тренировочную выборку. В задачах классификации при этом происходит неявное выделение ключевых признаков внутри сети, определение значимости признаков и системы взаимоотношений между ними. В настоящее время разработаны мощные, гибкие и универсальные механизмы обучения различных типов НС [20,22,29,69]. Кроме того, архитектура НС и процедуры обучения позволяют выполнить гибкую настройку на конкретную решаемую задачу. Для большинства НС процедура обучения является эвристическим алгоритмом, что, с одной стороны, обеспечивает приемлемость получаемых решений, а с другой стороны, не требует непомерных вычислительных ресурсов.

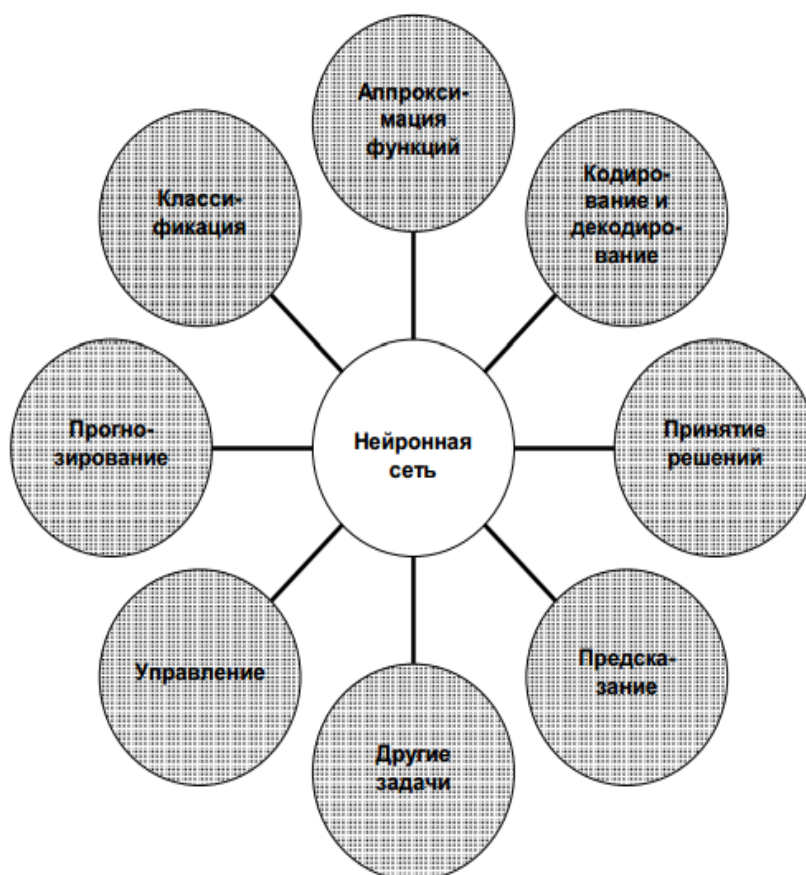


Рис. 3.6. Основные задачи решаемые с помощью нейронных сетей

Нейронные сети обладают хорошей обобщающей способностью (одной из лучших среди существующих методов, например, много лучшей, чем у решающих деревьев [20]). Это значит, что опыт, полученный в процессе обучения на конечном наборе образов, НС может успешно распространять на всё множество образов. Кроме интерполяционных обобщающих способностей, НС (многослойные персептроны, например) могут хорошо

экстраполировать, т.е. применять свой опыт на качественно иные образы, чем те, которые встречались в обобщающей выборке.

Нейронные сети ни налагают каких-либо ограничений на тренировочную выборку, ни полагаются на то, что она обладает какими-либо априорными свойствами, в отличие, например, от статистических методов [32,85]. Не требуется никакого предварительного изучения характера данных. НС принимает тренировочный набор «как есть» и учится производить правдоподобное решение, не претендуя на абсолютную истину, т.е. строится наилучшая не физическая модель [19], которая не является максимально точным соответствием реального процесса, но даёт приемлемую его аппроксимацию. Имеется ряд примеров, когда нейронные сети показывали себя лучше статистических методов [20]. Кроме того, в статистике не имеется аналогов некоторых нейросетевых методов [32], таких, например, как карты Кохонена, машина Больцмана и, что важно для распознавания изображений, когнитрон.

Естественным образом архитектура НС реализуется на параллельных вычислительных средствах: специализированных микросхемах, оптических и квантовых компьютерах. Это открывает широкие перспективы применения НС в будущем. НС характеризуется нечёткими распределённым хранением информации, т.е. нет отдельного нейрона, отвечающего за какое-либо понятие или признак, и удаление или искажение работы этого нейрона не приведёт к фатальным последствиям.

Но несмотря на все достоинства, применение НС к изображениям требует специальных усилий. Это связано в первую очередь со сложным характером изображений, особенно изображений трёхмерных объектов реального мира, какими и являются лица людей. Изображение должно быть предобработано – приведено к некоторым стандартным условиям. Кроме того, выбор начального представления изображения (это могут быть, например, частотные коэффициенты, главные компоненты, вейвлетные коэффициенты, моменты и т.п.) является отдельной обширной темой. Двумерный характер изображения, изменение условий освещённости, топологические искажения изображения при смене ракурса и прочих воздействиях не позволяют ограничиться простейшими архитектурами НС для достижения оптимального результата. Кроме различных способов применения нейронных сетей к распознаванию человека по изображению лица, существует множество работ, посвящённых применению нейронных сетей к распознаванию и обработке изображений других видов объектов.

3.1.6. Модель искусственного нейрона

Представление о детальном устройстве головного мозга появилось только около ста лет назад. В 1888 г. испанский доктор Рамони Кайал экспериментально показал, что мозговая ткань состоит из большого числа связанных друг с другом однотипных узлов – нейронов. Более поздние исследования при помощи электронного микроскопа показали, что все нейроны, независимо от типа, имеют схожую организационную структуру показанную на рис 3.7.[11,20,22,86] естественная нервная клетка (нейрон) состоит из тела (сомы), содержащего ядро, и отростков – дендритов, по которым в нейрон поступают входные сигналы. Один из отростков, ветвящийся на конце, служит для передачи выходных сигналов данного нейрона другим нервным клеткам. Он называется аксоном. Соединение аксона с дендритом другого нейрона называется синапсом. Нейрон возбуждается и передает сигнал через аксон, если число пришедших по дендритам возбуждающих сигналов больше, чем число тормозящих.

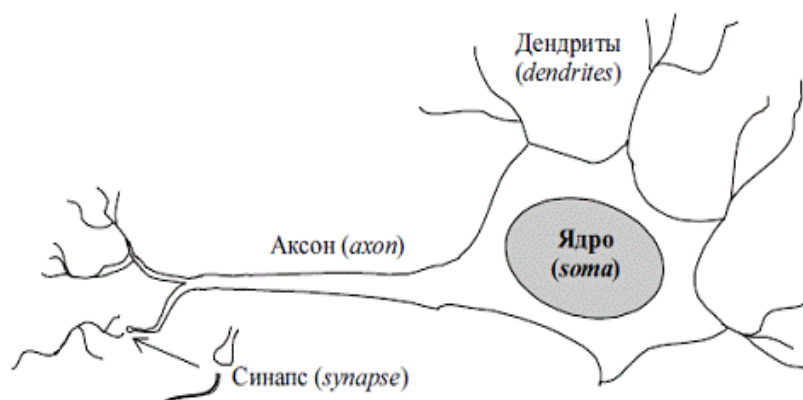


Рис. 3.7. Строение биологического нейрона.

В 1943 г. В. Маккаллох и В. Питтс предложили систему обработки информации в виде сети, состоящей из простых вычислителей, созданных по принципу биологического нейрона. Искусственная нейронная сеть (ИНС) представляет собой совокупность простых вычислительных элементов (процессоров) – искусственных нейронов (рис. 3.8), соединенных некоторым образом так, чтобы между ними обеспечивалось взаимодействие. Искусственные нейроны характеризуются правилом комбинирования входных сигналов и передаточной функцией, позволяющим вычислить выходной сигнал.

Информация, поступающая на вход нейрона, суммируется с учетом весовых коэффициентов сигналов:

$$P_j = \sum_{i=1}^d x_i w_i + w_0 \quad (3.1.)$$

где w_0 – сдвиг (порог, смещение) нейрона.

В зависимости от значения весового коэффициента w_i , входной сигнал x_i или усиливается, или подавляется. Взвешенную сумму входных сигналов также называют потенциалом или комбинированным входом нейрона.

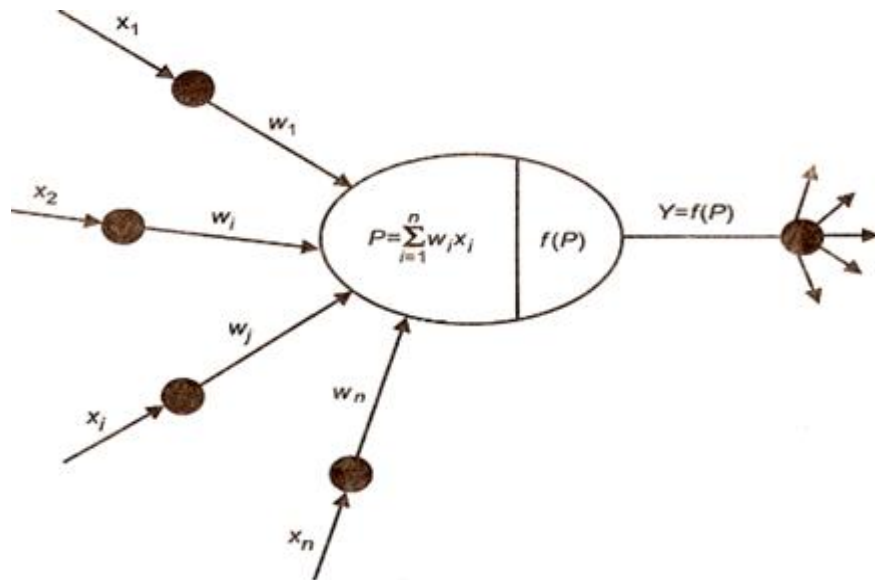


Рис. 3.8. Кибернетическая модель нейрона.

Сдвиг обычно интерпретируется как связь, исходящая от элемента, активность которого всегда равна 1. Обычно, для удобства входной вектор расширяется добавлением этого сигнала до $x = (1, x_0, \dots, x_n)$ и порог w_0 вносится под знак суммы:

$$P_j = \sum_{i=0}^d x_i w_i \quad (3.2.)$$

Передающая функция, или функция активации нейрона, является правилом, согласно которому взвешенная сумма поступивших сигналов P преобразуется в выходной сигнал нейрона Y , который передается другим нейронам сети, т. е. $Y=f(P)$. На рисунке 3.9. показаны графики наиболее распространённых функций активации нейронов.

Пороговая функция пропускает информацию только в том случае, если алгебраическая сумма входных сигналов превышает некоторую постоянную величину P^* , например:

$$Y = \begin{cases} 1, & P \geq P^* \\ -1, & P < P^* \end{cases} \quad (3.3.)$$

Пороговая функция не обеспечивает достаточной гибкости

ИНС при обучении. Если значение вычисленного потенциала не достигает заданного порога, то выходной сигнал не формируется и нейрон «не срабатывает». Это приводит к снижению интенсивности выходного сигнала нейрона и, как следствие, к формированию невысокого значения потенциала взвешенных входов в следующем слое нейронов.

Линейная функция дифференцируема и легко вычисляется, что в ряде случаев позволяет уменьшить ошибки выходных сигналов в сети, так как передаточная функция сети также является линейной. Однако она не универсальна и не обеспечивает решения многих задач.[23,38]

Определенным компромиссом между линейной и ступенчатой функциями является сигмоидальная функция активации

$$Y = 1/(1+exp(-kP)) \quad (3.4),$$

которая удачно моделирует передаточную характеристику биологического нейрона (рис. 3.9., в).

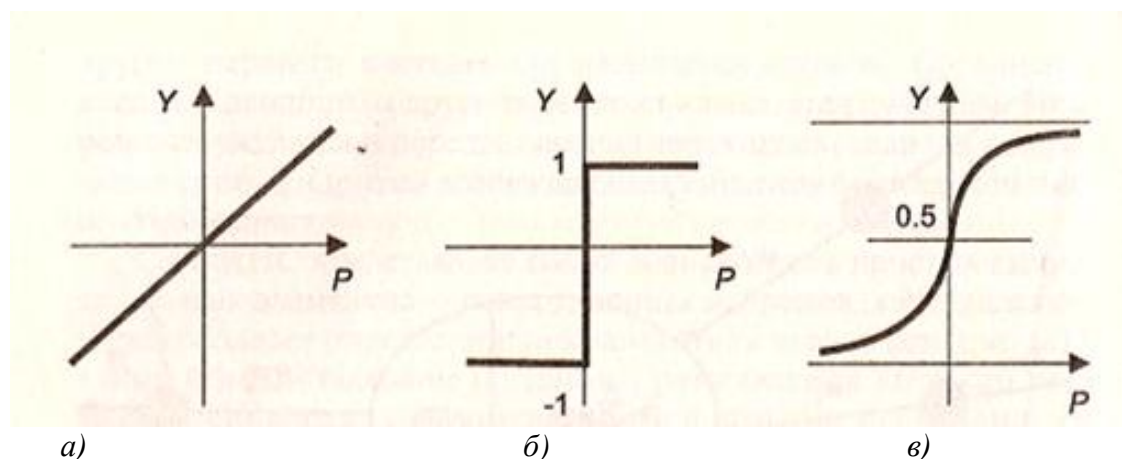


Рис. 3.9. Функции переноса искусственных нейронов:
 а) линейная; б) ступенчатая; в) сигмоидальная.

Коэффициент k определяет крутизну нелинейной функции: чем больше k, тем ближе сигмоидальная функция к пороговой; чем меньше k, тем она ближе к линейной. Тип функции переноса выбирается с учетом конкретной задачи, решаемой с применением нейронных сетей. Например, в задачах аппроксимации и классификации предпочтение отдают сигмоидальной кривой.

Архитектура ИНС. С каждым нейроном связывается набор входящих связей, по которым к данному элементу поступают сигналы от других элементов сети, и набор исходящих связей, по которым сигналы данного элемента передаются другим нейронам. Некоторые нейроны предназначены для получения сигналов из внешней среды (входные элементы), а некоторые – для вывода во внешнюю среду результатов вычислений (выходные элементы).[11,20]

В 1958 г. Фрэнк Розенблатт предложил следующую модель

нейронной сети – перцептрона. Перцептрон Розенблатта (рис. 3.10.) состоит из k нейронов, имеет d входов, k выходов и только один слой настраиваемых весов w_{ij} .

Входные нейроны обычно предназначены для распределения вводимых сигналов между другими нейронами сети, поэтому для них требуется, чтобы исходящий от элемента сигнал был таким же, как и входящий. В отличие от других нейронов сети, входные имеют только по одному входу. Иными словами, каждый входной элемент может получать сигнал от одного соответствующего ему датчика. Поскольку входные элементы предназначены исключительно для того, чтобы распределять сигналы, получаемые из внешней среды, многие исследователи вообще не считают входные элементы частью нейронной сети.

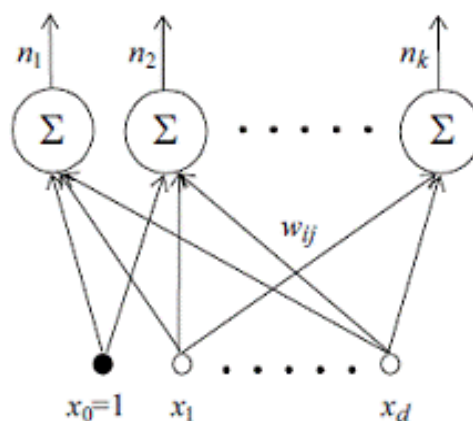


Рис. 3.10. Перцептрон Розенблатта.

Перцептрон способен решать линейные задачи. Число входов сети определяет размерность пространства, из которого выбираются входные данные: для двух признаков пространство оказывается двумерным, для трех – трехмерным, а для d признаков – d -мерным. Если прямая или гиперплоскость в пространстве входных данных может разделить все образцы на соответствующие им классы, то проблема является линейной, в противном случае – нелинейной. На рис. 3.11. показаны множества точек на плоскости, причём в случае а) граница линейная, в случае – б) нелинейная.

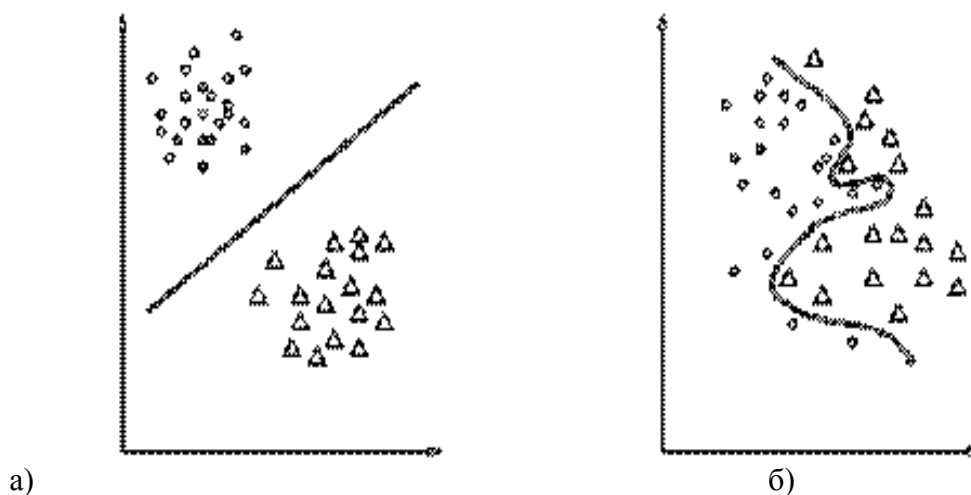


Рис. 3.11. Геометрическое представление линейной (а) и нелинейной (б) задач.

Для решения нелинейных проблем предложены модели многослойных перцептронов (MLP), способные строить ломаную границу между распознаваемыми образами. В многослойных сетях каждый нейрон может посылать выходной сигнал только в следующий слой и принимать входные сигналы только с предыдущего слоя, как показано на рис. 3.12. Слои нейронов, расположенные между входным и выходным называются скрытыми, так как не получают и не передают данные непосредственно из внешней среды. Такая сеть позволяет выделять глобальные свойства данных за счет наличия дополнительных синаптических связей и повышения уровня взаимодействия нейронов.

Определение числа скрытых слоев и числа нейронов в каждом слое для конкретной задачи является неформальной проблемой, при решении которой можно использовать эвристическое правило: число нейронов в следующем слое в два раза меньше, чем в предыдущем

3.1.7. Классификация нейронных сетей

В настоящее время кроме многослойного перцептрона существует множество способов задания структур нейронных сетей. Все виды нейронных сетей можно условно разделить на сети прямого распространения и сети с обратными связями. Как следует из названия, в сетях первого типа сигналы от нейрона к нейрону распространяются в четко заданном направлении – от входов сети к ее выходам. В сетях второго типа выходные значения любого нейрона сети могут передаваться к его же входам. Это позволяет нейронной сети моделировать более сложные процессы, например временные, но делает выходы подобной сети нестабильными, зависящими от состояния сети на предыдущем цикле. На рис. 3.13.

представлена такая классификация наиболее распространенных типов НС.[11,20].

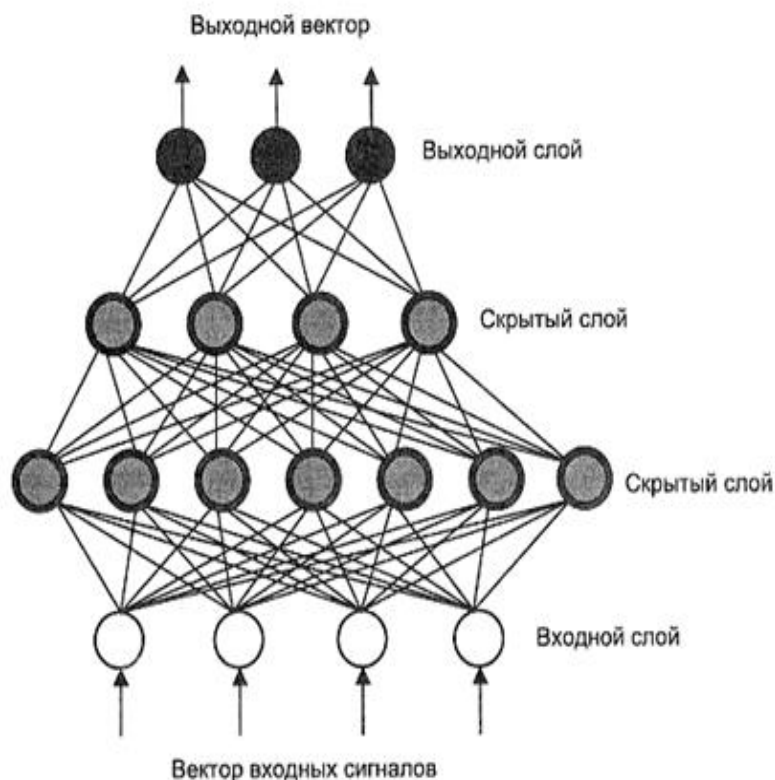


Рис. 3.12. Схема многослойного перцептрона.

По характеру связей нейронные сети могут быть полносвязными, когда каждый нейрон связан со всеми остальными, и слоистыми, когда нейроны последующего слоя связаны только со всеми нейронами предыдущего слоя. Эти две архитектуры являются базовыми, но возможны и различные вариации.

По характеру функционирования нейронные сети могут быть однопроходными, когда выход сети рассчитывается за один проход сети и релаксационными, когда функционирование сети продолжается до достижения стабильного состояния, это состояние и является результатом работы.

По характеру формирования связей нейронные сети могут быть следующих видов[23]:

Обучение с учителем: связи настраиваются в процессе обучения, причём эталонные значения результатов работы известны.

Самообучение (обучение без учителя): эталонные результаты неизвестны (не нужны), сеть в процессе обучения должна организовать входные образы на основе их подобия.

Фиксированные связи, определяемые характером решаемой задачи (например в оптимизационных задачах).

Так же нейронные сети могут отличаться типом входной

информации (двоичная, аналоговая и т.п.) и методом обучения.



Рис. 3.13. Классификация распространённых видов ИНС.

3.2. Разделение пространства признаков на области и извлечение ключевых признаков

Для решения данных задач предназначены многослойные нейронные сети, нейронные сети высокого порядка и радиально-базисные нейронные сети. Поскольку такие сети оперируют в исходном пространстве изображений (признаков), то для них является критическим требование предобработки изображения. Это приведение изображения к стандартному виду (положение, масштаб, ориентация, выравнивание яркости), снижение размерности данных, выбор ключевых характеристик. Следующим следствием оперирования в исходном пространстве является невозможность учёта искажения изображения (например, при изменении ракурса, эмоций), и поэтому тренировочная выборка должна содержать репрезентативный набор примеров, представляющих собой наборы изображений объектов в том диапазоне ракурсов и условий освещения, в которых планируется применение системы распознавания [18].

3.2.1. Многослойные нейронные сети

Архитектура многослойной нейронной сети (МНС, другое название.— многослойный персептрон, по-английски Multilayer Perceptron, MLP) состоит из последовательно соединённых слоёв, где

нейрон каждого слоя своими входами связан со всеми нейронами предыдущего слоя, а выходами – следующего (рис. 4314). Активационными функциями для таких нейронов служат разновидности линейных, пороговых и сигмоидных функций [20].

НС с одним решающим слоем способна формировать линейные разделяющие поверхности, что сильно сужает круг решаемых задач, в частности, такая сеть не сможет решить задачу типа «исключающее или». НС с нелинейной функцией активации и двумя решающими слоями позволяет формировать любые выпуклые области в пространстве решений, а с тремя решающими слоями – области любой сложности, в том числе и невыпуклой формы [20]. При этом МНС не теряет своей обобщающей способности. С помощью двухслойной НС можно с любой точностью аппроксимировать любую многомерную функцию на отрезке от 0 до 1. Обучаются МНС при помощи алгоритма обратного распространения ошибки, являющегося разновидностью градиентного спуска в пространстве весов с целью минимизации суммарной ошибки сети:

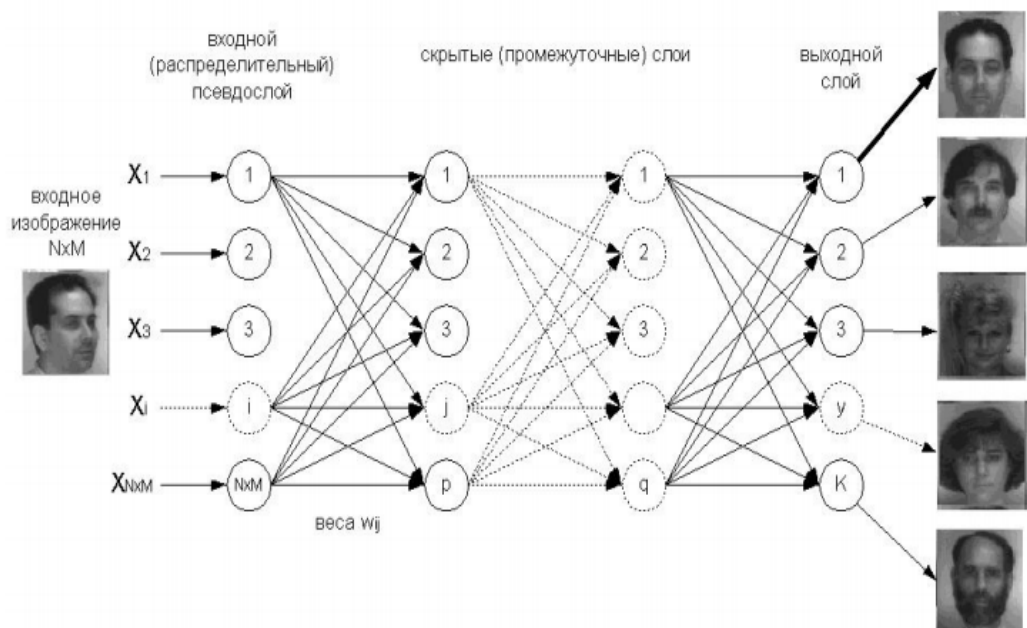


Рис. 3.14. Архитектура многослойной нейронной сети и её применение для распознавания изображений. Нейрон с максимальной активностью (здесь первый) указывает принадлежность к распознанному классу

$$\Delta W = -\alpha \frac{dE}{dW}, \quad E = \frac{1}{2} \sum_j (y_j - t_j)^2, \quad (3.5.)$$

где t_j – эталонное значение выходов сети. При этом ошибки (точнее величины коррекции весов) распространяется в обратном

направлении от входов к выходам, сквозь веса, соединяющие нейроны. Алгоритм обратного распространения ошибки является *NP*-трудным, поэтому время обучения сети увеличивается экспоненциально с ростом размерности данных. Так как эталонные значения выходов известны, алгоритм является обучением с учителем. Но в применении к извлечению ключевых признаков, когда рециркуляционная сеть обучается реконструировать поданное на вход изображение, а на скрытых нейронах формируется его сжатое представление, алгоритм обучения может быть назван и самообучением. МНС, как и большинство других типов НС, перед началом обучения инициализируется случайными весами. Поэтому две разные обученные НС, имеющие одинаковый показатель ошибки, часто представлены совершенно различными разделяющими поверхностями, несводимыми друг к другу. На этом основан метод коллективов (ансамблей) нейронных сетей, часто применяемый в распознавании человека по изображению лица [18,33].

Суть его заключается в том, что имеется набор (коллектив) сетей, обученных решать одну и ту же задачу, но различными способами (различная начальная инициализация весов, архитектура, порядок следования примеров при обучении и т.п.). Обобщённое решение такого коллектива как правило точнее (и надёжнее), чем решение единственной нейронной сети. Существуют так же методики синтеза коллектива сетей, производящих максимально независимые ошибки [12].

Другое направление развития архитектуры МНС – это нейронные деревья [11,20]. В этой архитектуре узлами решающего дерева являются нейронные сети. По мере продвижения от корня дерева сети-узлы уточняют решение задачи. В работе [70] был разработан новый алгоритм создания и обучения нейронных деревьев, который применялся для классификации сцен на изображении. По сравнению с такими методами, как обычные нейронные деревья, решающие деревья, коллективы НС и многослойные нейронные сети, точность распознавания у предложенного алгоритма была сравнима или выше, а скорость обучения и работы – на порядок выше. Однако обобщающая способность такой архитектуры в [70] не проверялась.

Одними из главных проблем МНС (и некоторых других типов НС) являются следующие [3,6,9,10,21]:

1. Проблема локального минимума. Как и для всех градиентных методов, проблема локального минимума заключается в том, что при итерационном спуске может наступить момент, когда решение заходит в локальный минимум, из которого вследствие малой величины шага не может выбраться. И такой локальный минимум не всегда обеспечивает приемлемое решение. Выход

заключается в применении стохастических методов [12,20].

2. Выбор архитектуры сети (количество нейронов, слоёв, характер связей). С этим также связана проблема переобучения, которая заключается в том, что сеть с избыточным числом элементов теряет обобщающую способность и хорошо работает только на тренировочной выборке. В настоящее время разработаны различные априорные оценки выбора архитектуры, методы прореживания обученных сетей, методы «растущих» сетей [11,12].

3. Выбор шага (скорости) обучения. Такая проблема связана с тем, что при малом шаге время обучения будет большим и сеть может застревать в локальных минимумах, а при больших шагах возможно расхождение процесса обучения или паралич сети. Проблема эффективно решается адаптивным шагом [12], который для каждой итерации позволяет сделать шаг, минимизирующий ошибку сети на данной итерации. Существуют методы, которые на каждом тренировочном цикле (называемом эпохой) анализируют всю тренировочную выборку и выбирают оптимальное значение и направление шага [12].

Одним из наиболее перспективных методов, применяемых на этапе обучения НС, является генетический алгоритм (ГА, по-английски – Genetic Algorithm, GA), относящийся к эволюционным методам [6]. Генетический алгоритм представляет собой параллельный асинхронный оптимизационный метод [6]. Поиск решения в нём осуществляется одновременно целой популяцией хромосом (хромосома – единичное закодированное решение задачи). Хромосомы могут как обмениваться опытом друг с другом, улучшая свою пригодность (оператор скрещивания), так и осваивать новые области решения (оператор мутации). Оптимизационный критерий задаёт функцию оценки пригодности хромосом, и процесс эволюции популяции представляет собой улучшение решения задачи. Главным преимуществом ГА является то, что при линейном увеличении размера популяции, скорость поиска решения возрастает экспоненциально (т.н. скрытый параллелизм) [6,12]. Это позволяет достичь лучших субоптимальных решений.

Поскольку обучение сети – это оптимизационный процесс, генетический алгоритм естественным образом вкладывается в алгоритм обучения НС [6]. При этом для ускорения ГА может быть введён оператор локального (градиентного) спуска, представляющий собой метод обратного распространения в применении к отдельной хромосоме-сети. Так же, закодировав в хромосому структуру сети, и введя в оценочный критерий ГА штраф за избыточность архитектуры, можно добиться синтеза НС с минимально-необходимой архитектурой, обладающей хорошими обобщающими способностями [19].

Рассмотрим применение многослойных нейронных сетей к распознаванию человека по изображению лица.

Простейшее применение однослойной НС [11] (называемой автоассоциативной памятью) заключается в обучении сети восстанавливать подаваемые изображения. Подавая на вход неизвестное изображение и вычисляя качество реконструированного изображения, можно оценить, насколько сеть распознала входное изображение. Положительные свойства этого метода заключаются в том, что сеть может восстанавливать искажённые и зашумленные изображения, но для более серьёзных целей он не подходит.

МНС также используется для непосредственной классификации изображений – на вход подаётся или само изображение в каком-либо виде, или набор ранее извлечённых ключевых признаков изображения, на выходе нейрон с максимальной активностью указывает принадлежность к распознанному классу (см. рис.3.12). Если эта активность ниже некоторого порога, то считается, что поданный образ не относится ни к одному из известных классов. Процесс обучения устанавливает соответствие подаваемых на вход образов с принадлежностью к определённому классу. Это называется обучением с учителем. В экспериментах на тестовой базе ORL такой подход позволил достичь стабильной 93%-ной точности распознавания (98%-ной максимальной). В применении к распознаванию человека по изображению лица такой подход хорош для задач контроля доступа небольшой группы лиц. Он обеспечивает непосредственное сравнение сетью самих образов, но с увеличением числа классов время обучения и работы сети возрастает экспоненциально. Поэтому такие задачи, как поиск похожего человека в большой базе данных, требуют извлечения компактного набора ключевых характеристик, на основе которых можно производить поиск.

В работе [82] для классификации изображений лиц использовался многослойный персептрон. В качестве входных данных применялись результаты преобразования симметрии. При использовании коллектива нейронных сетей точность распознавания повышалась. Точность повышалась и при предварительной обработке входных данных самоорганизующимися сетями. Всё это позволило достичь 0.62% ошибки на базе данных MIT.

В работе [51] для классификации так же использовалась многослойная нейронная сеть, но входными данными являлось изображение, подвергнутое многомерному снижению размерности, и результаты применения различных вариаций Гауссовских фильтров. Такой подход позволил достичь высокой обобщающей способности, имея в обучающей выборке только одно изображение на каждого человека.

Подход к классификации с использованием частотных характеристик всего изображения описан в [85]. Применялась однослойная НС, основанная на многозначных нейронах, которая, как и многослойный персептрон, способ настроить сложные разделяющие поверхности, но по иному принципу. Спектральные коэффициенты преобразования Фурье выбирались по правилу «зигзага», и результат стабилизировался при числе первых коэффициентов начиная с 405. Отмечено 100%-ное распознавание на базе данных MIT.

В работах [59,68] использовались первые 25-35 коэффициентов дискретного косинусного преобразования (из 10304, т.е. около 0.2% от общего количества коэффициентов) и двухслойная нейронная сеть. Коэффициенты выбирались двумя способами. В первом способе брались первые коэффициенты из левого верхнего угла матрицы коэффициентов, ограниченного диагональю. Во втором способе выбирались коэффициенты, которые давали наименьшую погрешность реконструкции изображения. Разница в точности при этом была неразличима. И хотя реконструкция по таким коэффициентам невозможна (рис. 3.15), точность распознавания составила в среднем 96% (на базе ORL). Использовалось 60 и 80 нейронов в скрытом слое (всего два слоя, в выходном – 40 нейронов, по числу классов).



Рис. 3.15. Реконструкция по первым 28 из 10304 коэффициентов, изображение 92x112

Применение МНС для классификации изображений лиц на основе таких характеристик, как расстояния между некоторыми специфическими областями лица (нос, рот, глаза), описано в [59]. В этом случае на вход НС подавались эти расстояния. Использовались также комбинированные методы: в первом из них на вход НС подавались результаты обработки скрытой Марковской моделью, а во втором – результат работы НС подавался на вход скрытой Марковской модели. Во втором случае преимуществ не наблюдалось.

Это говорит о том, что скрытые Марковские модели сами по себе обладают худшей различающей способностью, чем НС.

В работе [51] показано применение НС для классификации изображений, когда на вход сети поступают результаты декомпозиции изображения по методу главных компонент. Классификация главных компонент при помощи НС повышала точность распознавания. МНС применяются и для обнаружения объектов определённого типа. Помимо того что любая обученная МНС в некоторой мере может определять принадлежность образов к «своим» классам, её можно специально обучить надёжному детектированию определённых классов. В этом случае выходными будут два класса: класс принадлежащих и класс не принадлежащих к заданному типу образов. В [99] применялся нейросетевой детектор для обнаружения изображения лица во входном изображении. Изображение сканировалось скользящим окном 20x20 пикселей, которое подавалось на вход сети, решающей, принадлежит ли данный участок к классу лиц. Обучение производилось с использованием как положительных примеров (различных изображений лиц), так и отрицательных (изображений, не являющихся лицами). В обучающей выборке незначительно варьировались масштаб, положение и ориентация лица. Для повышения надёжности обнаружения использовался коллектив НС, обученных с различными начальными весами, вследствие чего НС ошибались по-разному, а окончательное решение принималось голосованием всего коллектива.

НС применяется также для извлечения ключевых характеристик изображения, которые затем используются для последующей классификации. В [12,101], показан способ нейросетевой реализации метода главных компонент. Суть метода главных компонент заключается в получении максимально декоррелированных коэффициентов, характеризующих входные образы. Такие коэффициенты называются главными компонентами и используются для статистического сжатия и реконструкции изображений. При этом небольшое число коэффициентов используется для представления всего образа. Каждое изображение разлагается на линейную комбинацию собственных векторов. Для набора изображений лиц собственные векторы могут быть представлены в виде изображений, такие изображения похожи на лица и называются собственными лицами (eigenfaces, рис. 3.16). Сумма собственных векторов, умноженных на соответствующие им главные компоненты, представляет собой реконструкцию изображения.

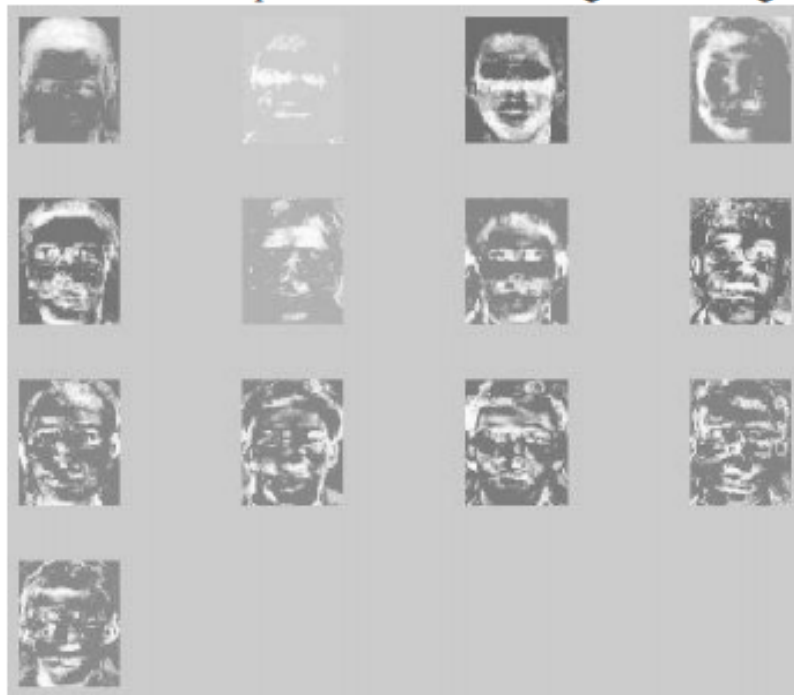


Рис. 3.16. Собственные лица (eigenfaces)

НС с одним скрытым слоем, содержащим m нейронов, число которых много меньше, чем размерность изображения $m \ll n$, обученная по методу обратного распространения ошибки восстанавливать на выходе изображение, поданное на вход, формирует на выходе скрытых нейронов коэффициенты первых m главных компонент, которые и используются для сравнения изображений. Архитектура такой сети, называемой рециркуляционной нейронной сетью (РНС), показана на рис. 3.17.

Обычно используется от 10 до 200 главных компонент. С увеличением номера компоненты её репрезентативность сильно понижается, и использовать компоненты с большими номерами не имеет смысла. Использование линейных активационных функций в НС позволяет получить на выходе скрытого слоя именно m первых главных компонент, аналогичных получаемым при решении матричных уравнений. При использовании нелинейных активационных функций нейронных элементов возможна нелинейная декомпозиция на главные компоненты. Нелинейность позволяет более точно отразить вариации входных данных, однако при этом выходы скрытых нейронов будут только похожи на главные компоненты. Веса, сформировавшиеся при таком обучении на входном и выходном слое, так же будут похожи на собственные лица [80,102], которым присуще полезное свойство – существуют компоненты, которые в основном отражают такие существенные характеристики лица, как пол, раса, эмоции. Первые компоненты отражают наиболее общую форму лица, последние – различные

мелкие отличия между лицами. Такой метод хорошо применим для поиска похожих изображений лиц в больших базах данных.

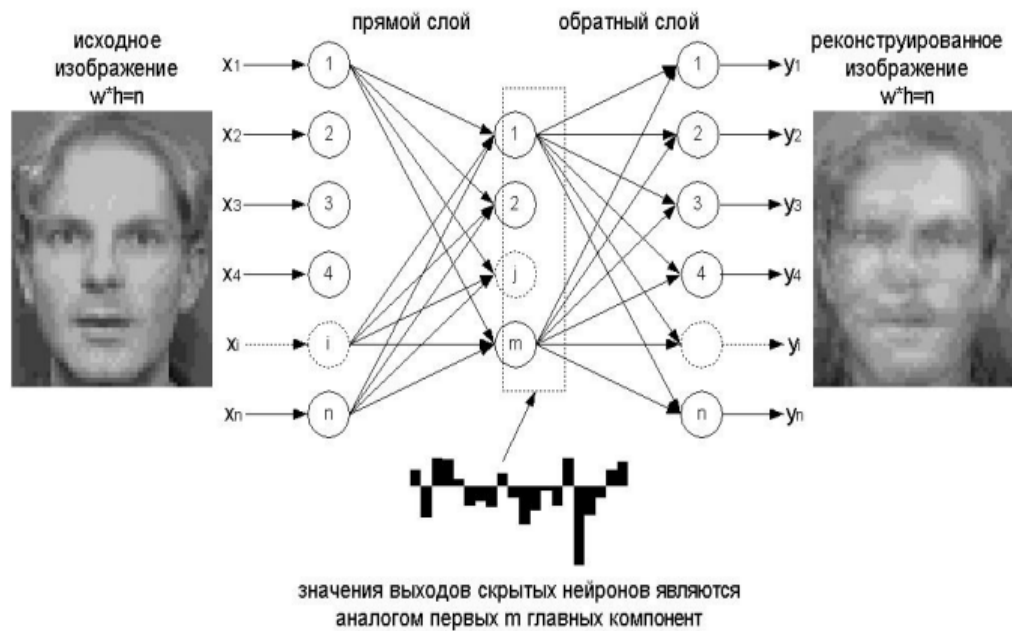


Рис. 3.17. Архитектура рекуррентной нейронной сети для извлечения главных компонент

Этот метод также используется в задаче обнаружения лица на изображении. Оценивая качество реконструкции входного изображения, можно очень точно определять его принадлежность к классу лиц. Для изображений, не являющихся лицами, реконструкция будет невысокого качества.

В работе [101] использовалась МНС для дальнейшей классификации главных компонент. Производилась идентификация человека по изображению лица, определение его расы, пола, эмоционального состояния.

Преимущества применения РНС для извлечения главных компонент перед решением матричных уравнений [52]:

- -алгоритм обучения РНС прост и универсален;
- -нелинейная активационная функция позволяет точнее реконструировать изображение;
- -при решении матричных уравнений возможны проблемы, если примеры очень похожи друг на друга, РНС лишена такого недостатка;
- -не требуется вычислять все собственные векторы. Таким образом, время обучения сети линейно зависит от количества извлекаемых главных компонент;
- -для предварительных экспериментов можно использовать меньшее число обучающих циклов, что снижает время обучения.

В работе [52] показана возможность дальнейшего уменьшения размерности главных компонент при помощи НС. Суть её заключается в использовании большего количества скрытых слоёв,

отвечающих за сжатие и реконструкцию изображения. Такое сжатие позволяет уловить более сложные закономерности в наборе образов и, следовательно, представить их точнее и меньшим числом компонент. Заметим, что такой подход может быть использован и для решения некоторых других задач, например, для выделения (детекции) участков кожи (областей лица и кистей рук человека на изображении) для автоматизации распознавания жестовой речи [41].

3.2.2. Нейронные сети высокого порядка и моментные НС

Нейронные сети высокого порядка (НСВП, по-английски – High Order Neural Network) отличаются от МНС тем, что у них только один слой, но на входы нейронов поступают так же термы высокого порядка, являющиеся произведением двух или более компонент входного вектора[11], например, для сетей второго порядка:

$$S = \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j - T. \quad (3.6)$$

Такие сети так же могут формировать сложные разделяющие поверхности. Разделяющая поверхность второго порядка $S = 0$ называется гиперквадрикой[10]. Добавляя компоненты входного вектора в произведение, получим класс полиномиальных разделяющих поверхностей. Такие сети также можно обучать по методу обратного распространения. Многослойные НС в общем случае эффективнее, но существует ряд приложений, в которых сети высокого порядка лучше чем МНС.

В работе [8] описаны перспективные архитектуры и методы обучения нейронных сетей высокого порядка и комбинированных нейронных сетей для распознавания изображений инвариантно к сдвигу, масштабу и повороту. Архитектура таких сетей высокого порядка основана на вычислении моментов изображения первыми слоями.

В работе [50] показано применение НСВП третьего порядка (рис. 3.18) для распознавания изображений лиц, имеющих произвольные масштаб и ориентацию изображения. Приведены методы обучения такой сети. Особенность её заключается в том, что для обучения некоторому классу достаточно предъявить его образ без вариаций масштабов и поворотов – после обучения сеть будет распознавать известные классы инвариантно к масштабу и поворотам изображения. Такая сеть не является полносвязной, быстро обучается и работает. Отмечено существенное повышение точности классификации такой сетью повернутых и масштабированных изображений по сравнению с МНС.

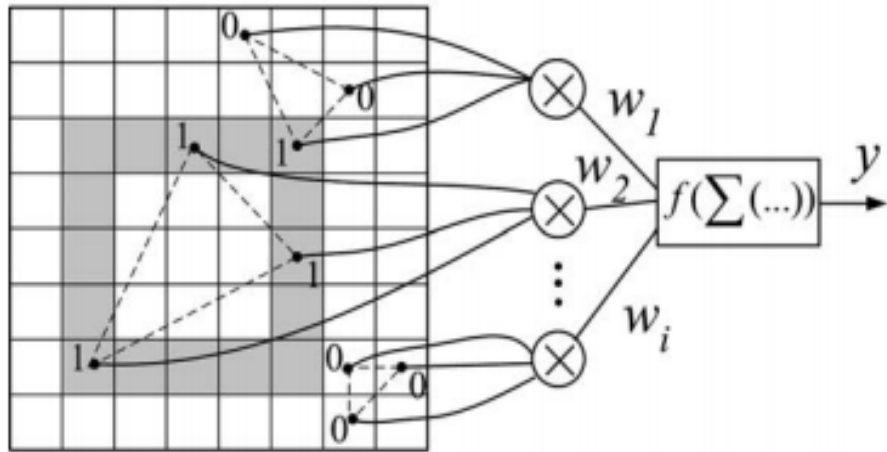


Рис. 3.18. НСВП третьего порядка

3.2.3. Радиально-базисные нейронные сети

Радиально-базисные нейронные сети (РБНС, по-английски – Radial Basis Function Network, RBF) состоят из двух слоёв, рис. 3.19[6].

Первый слой имеет радиально-базисную активационную функцию:

$$y = \exp\left(\frac{-S^2}{2\sigma^2}\right), \quad (3.7.)$$

где σ – среднеквадратичное отклонение, характеризующее ширину функции

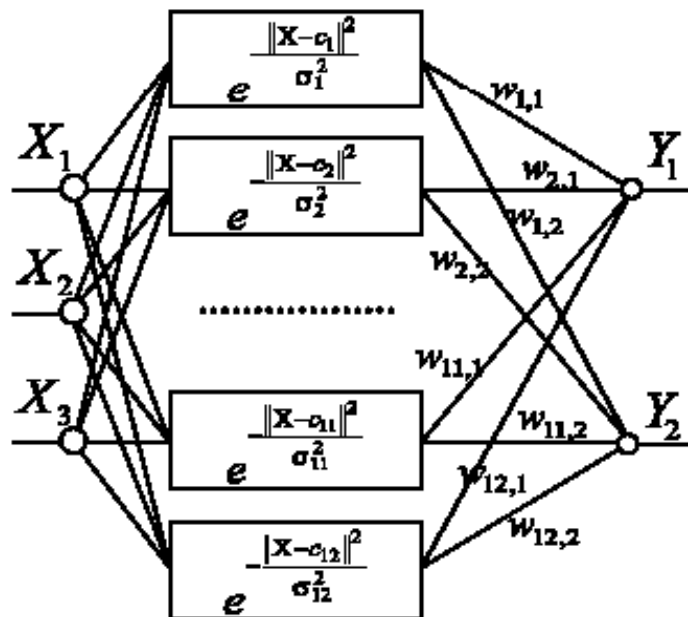


Рис. 3.19. Радиально-базисная нейронная сеть, X_i – элементы входного вектора, Y_i – элементы выходного вектора, квадратами обозначены радиально-базисные нейроны

(размер кластера), S определяется как расстояние между входным и весовым вектором:

$$S^2 = \|X - W\|^2 = \sum_i (x_i - w_i)^2 \quad (3.8),$$

являющимся по сути, расстоянием до центра кластера, определяемым конкретным нейроном. Таким образом, скрытый слой представляет собой набор кластеров в пространстве образов и выполняет первый этап кластеризации входного образа – значение активационной функции каждого нейрона быстро уменьшается с удалением от центра кластера. Второй слой нейронов имеет линейную активационную функцию, и выполняет второй этап кластеризации – распределяет кластеры по классам. В отличие от карт Кохонена здесь обнуление нейронов, не обладающих максимальным выходным значением не требуется, они все вносят вклад в классификацию, и это преимущество РБНС.

РБНС также способна строить сложные разделяющие области и аппроксимировать многомерные функции. По сравнению с многослойной нейронной сетью, радиально-базисная сеть обучается на порядок быстрее, однако обладает намного худшей экстраполирующей способностью, т.е. не способна работать на образах, лежащих далеко от образов-примеров. Размеры РБНС больше, чем МНС для аналогичных задач, и РБНС становятся малоэффективны с ростом размерности входных данных [52].

Обучается такая сеть в два этапа. Первый этап осуществляется без учителя, на нём первый слой выделяет компактно расположенные группы кластеров. При этом корректируются центры кластеров. В настоящее время разработаны эффективные алгоритмы, позволяющие также подбирать оптимальный размер кластеров для каждого нейрона [6,29] и получать оптимальное количество нейронов в первом слое [6]. На втором этапе обучения второй слой учится распределять входные образы, пропущенные через первый слой, по классам. Информация об эталонных значениях выходов известна, обучение выполняется с учителем. Такое обучение производится или матричными методами, или алгоритмом обратного распространения ошибки [6,12,14].

В работе [67] применялись два способа извлечения ключевых характеристик и две различные архитектуры РБНС для распознавания лиц. В первом способе характеристики представляли собой набор главных компонент, во втором – коэффициенты вейвлет-преобразований. В первой архитектуре количество выходов соответствовало количеству классов, во второй применялся коллектив сетей, каждая из которых была обучена распознавать только свой класс. Отмечены значительные преимущества классификации РБНС перед непосредственным сравнением ключевых характеристик на

тестовой базе MIT.

В работе [67] применялись две различные архитектуры ансамблей РБНС для предварительной классификации изображений. На вход сети поступало изображение целиком, на выходах формировалась промежуточная классификация, которая затем подавалась на решающие деревья для контекстно-ориентированного распознавания изображений лиц (например: «найти все изображения определённого человека, где он в очках»). Различные сети в ансамблях первой архитектуры учились классифицировать изображения с различными типами изменений, второй – с одинаковыми, но количество нейронов менялось в процессе обучения. Решающий вывод делал «судья» (нейронная сеть, обученная обобщать решения коллектива НС), который принимал решение на основе голосования ансамбля сетей.

В работе [51] радиально-базисные сети использовались для распознавания человека по набору геометрических характеристик и определения его пола. Исходными служили полутоновые и рисованные (карикатурные) изображения. Отмечена хорошая способность РБНС выделять отличительные признаки.

В работе [55] использовалась РБНС совместно с оптическим потоком для анализа эмоционального выражения лица.

3.3. Топологически упорядоченное преобразование пространства

Векторные квантователи и самоорганизующиеся карты Кохонена (Self-Organizing Maps, SOM) используются для сжатия данных и извлечения ключевых характеристик [12]. Так же они служат основой для радиально-базисных сетей и когнитронов. Существуют, однако, и подходы, использующие карты Кохонена для решения оптимизационных задач (например, задачи коммивояжера), что не исключает их дальнейшего применения для распознавания изображений.

Векторные квантователи решают задачу квантования и кластеризации данных [12]. Архитектура таких сетей состоит из одного слоя (не считая входного распределительного), не имеет межнейронных связей, и является простейшим вариантом карт Кохонена. Входное пространство оптимальным образом разбивается на области-кластеры. В процессе обучения происходит выделение областей в исходном пространстве, каждой из которых соответствует отдельный нейрон. Такие сети функционируют по принципу «победитель берёт всё», активным считается нейрон, имеющий наибольшее выходное значение:

$$S_j = \sum_i w_{ij} x_i = W_j^T X \quad , (3.9.)$$

т.е. проекция входного вектора на веса которого оказалась максимальной. Выходом сети является номер нейрона-победителя.

Обучаются такие сети без учителя, т.е. они сами в процессе обучения выбирают оптимальное разбиение на области. Метод обучения таких сетей называется конкурентным. После подачи обучающего вектора выбирается нейрон, имеющий максимальную активность. Веса такого нейрона изменяются в сторону соответствия входному вектору, например:

$$W_j(t+1) = W_j(t) + \gamma(t)(X - W_j(t)) \quad . \quad (3.10)$$

Перед обучением веса инициализируются случайными значениями. Обучение ведётся до достижения стабилизации весов или завершается после определённого числа итераций.

Для того чтобы избежать ситуации, когда некоторые нейроны никогда не могут стать победителями и для более полного покрытия пространства (например в областях где плотность образов высока, требуется большее число нейронов) используются несколько подходов. Во первых, можно модифицировать веса проигравших нейронов с намного меньшей скоростью. Во вторых, можно вести статистику побед для каждого нейрона, и уменьшать возможности модификации слишком частых победителей.

Для автоматизации определения числа нейронов сети, используют такие алгоритмы, как например растущий нейронный газ. В нём нейроны, имеющие слишком большое число примеров или размер ячейки делятся на два нейрона[20].

Самоорганизующиеся карты Кохонена [11,20] позволяют получить топологически упорядоченное преобразование исходного n -мерного пространства в выходное m -мерное, $m \ll n$. Архитектура такой сети так же состоит из одного слоя, но нейроны в этом слое организованы в m -мерную решётку, и каждый нейрон имеет свою координату, определяющую его положение в решётке. Выходом сети являются координаты нейрона-победителя.

В правило модификации весов

$$W_j(t+1) = W_j(t) + \gamma(t)h(t,i,j)(X - W_j(t)) \quad (3.11)$$

вводится функция соседства h , убывающая с расстоянием между нейронами:

$$h(t,i,j) = \exp\left(\frac{-|i-j|^2}{2\sigma^2(t)}\right), \quad (3.12)$$

где i – нейрон-победитель, j – модифицируемый нейрон, $|i-j|$ –

расстояние между ними, $\sigma(t)$ – радиус области соседства, убывает со временем в процессе обучения. Для ускорения обучения используются растущие сети, когда заново создаваемые области решётки инициализируются значениями соседних нейронов.

Таким образом, обученная сеть способна топологически упорядоченно отображать входное пространство в выходное – векторы, близкие в исходном пространстве, будут иметь близкие координаты нейронов в решётке.

Это является особенно полезным при классификации данных, имеющих большое количество классов. Например, при классификации локальных участков изображений, может быть очень большое число классов, в которых переход от одного класса к другому практически непрерывен, затрудняя определение границ классов.

В работе [82] трёхмерная карта Кохонена (по пять узлов на каждое измерение) применялась для уменьшения размерности локальных участков 5×5 (размерность 25) изображений лиц (рис. 3.20). Каждому участку изображения 5×5 соответствует своя координата в карте Кохонена.

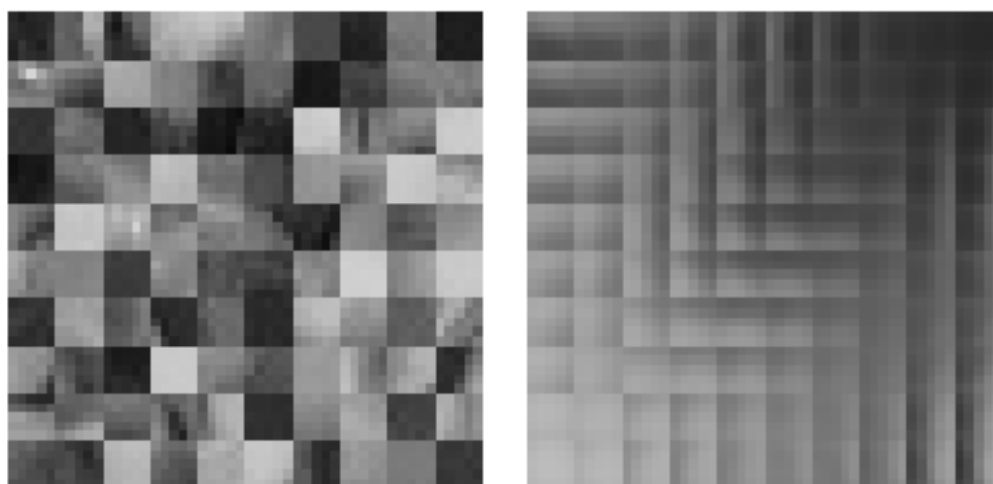


Рис. 3.20. Применение карт Кохонена для уменьшения размерности участков изображений лиц. Слева – топографическая карта участков изображений до, справа – после обучения.

После обучения похожие участки имеют близкое положение на карте. Входное изображение отображается на один из 125 узлов, положение которого в трёхмерной решётке кодирует вектор выходного пространства. Три измерения карты используются в качестве осей трёх ключевых характеристик (features) [100]. Такое преобразование обеспечивает частичную устойчивость к изменению освещения, смещениям и искажениям, избавляет от необходимости предварительной обработки изображения (преимущество – ускорение

работы), а так же значительно ускоряет процесс обучения и классификации, делая эту систему применимой в системах, работающих в реальном масштабе времени. В этой работе значения выходов карт Кохонена использовались для дальнейшего распознавания по изображениям лиц. Отмечено небольшое преимущество карт Кохонена перед методом главных компонент, которое заключалось в более высокой точности последующей классификации на основе данных уменьшенной размерности.

Карты Кохонена также применялись для обнаружения глаз на изображении лица [100]. Карта обучалась на типичных примерах изображений глаз. Наличие глаза в анализируемом участке изображения определялось по карте активности всех нейронов, в этом случае на ней наблюдались характерные пики, как показано на рис.3.21:

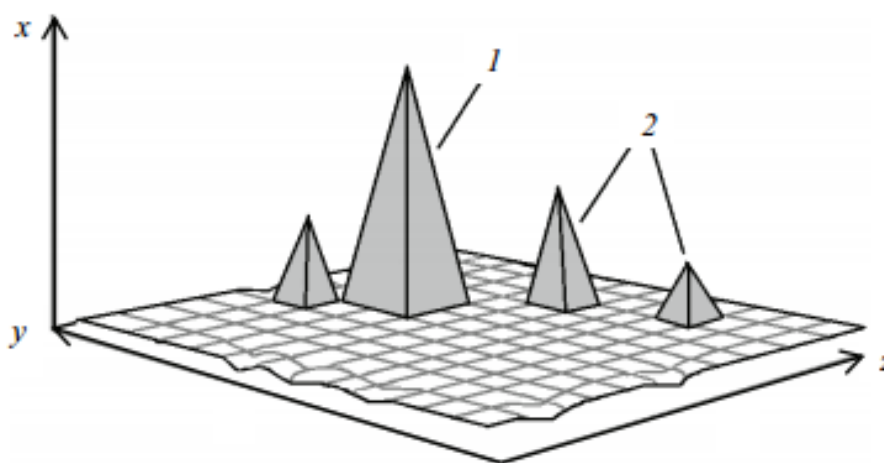


Рис. 3.21. Отклики нейронной сети. 1 – наиболее вероятное положение глаза, 2 – менее вероятные места расположения (шумы)

3.4. Распознавание с учётом топологии пространства

Описываемые далее типы нейронных сетей позволяют учесть топологию пространства изображения. Ведь изображение – это не просто n -мерный вектор, составленный из яркостей пикселей. Изображение имеет свою топологию, определяемую через двумерное локальное соседство пикселей [13].

Принципы работы таких сетей основываются на разбиении изображения на маленькие участки и иерархическом сопоставлении как взаимного их расположения, так и содержания. Такие сети являются наиболее перспективными для распознавания изображений.

3.4.1. Когнитрон

Когнитрон был разработан основываясь на анатомии и физиологии мозга, и своей архитектурой похож на строение зрительной коры [42]. Каждый слой мозга реализует различные уровни обобщения: входной слой чувствителен к простым образам, таким, как линии, и их ориентации в определенных областях визуальной области, в то время как реакция других слоев является более сложной, абстрактной и независимой от позиции образа. Аналогичные функции реализованы в когнитроне путем моделирования организации зрительной коры.

Главные архитектурные отличия когнитрона заключаются в том (рис. 3.22), что каждый нейрон связан только с небольшой локальной областью предыдущего слоя, и такие области перекрываются друг с другом. Слоев в когнитроне обычно больше чем в сетях других типов. Таким образом достигается иерархическая организация, когда на высших слоях когнитрон реагирует на более абстрактные образы, меньше реагирует на их смещение и искажение. Обучается когнитрон конкурентным обучением (без учителя).

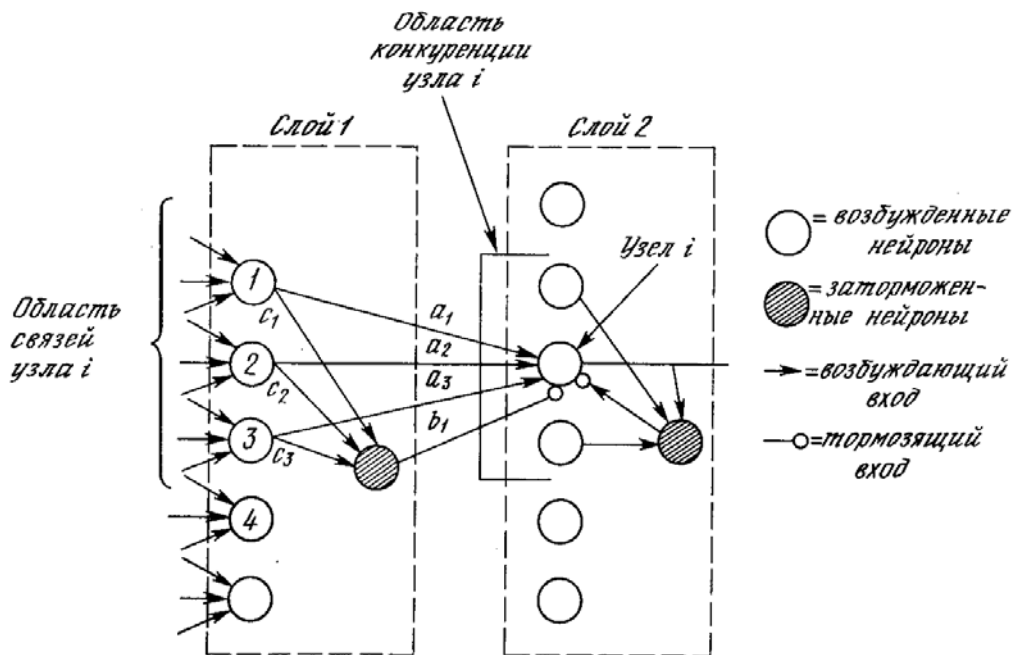


Рис. 3.22. Когнитрон

3.4.2. Неокогнитрон

В зрительной коре были обнаружены узлы, реагирующие на такие элементы, как линии и углы определенной ориентации [42]. На более высоких уровнях узлы реагируют на более сложные

и абстрактные образы такие, как окружности, треугольники и прямоугольники. На еще более высоких уровнях степень абстракции возрастает до тех пор, пока не определяются узлы, реагирующие на лица и сложные формы. В общем случае узлы на более высоких уровнях получают вход от группы низкоуровневых узлов и, следовательно, реагируют на более широкую область визуального поля. Реакции узлов более высокого уровня менее зависят от позиции и более устойчивы к искажениям.

Неокогнитрон является дальнейшим развитием идеи когнитрона и более точно отражает строение зрительной системы, позволяет распознавать образы независимо от их преобразований: смещения, вращения, изменения масштаба и искажения [42]. Неокогнитрон может как самообучаться, так и обучаться с учителем. Неокогнитрон получает на входе двумерные образы, аналогичные изображениям на сетчатой оболочке глаза, и обрабатывает их в последующих слоях аналогично тому, как это было обнаружено в зрительной коре человека.

Главное отличие неокогнитрона от когнитрона – это двумерная организация локальных участков и плоскостная иерархическая структура, рис. 3.23

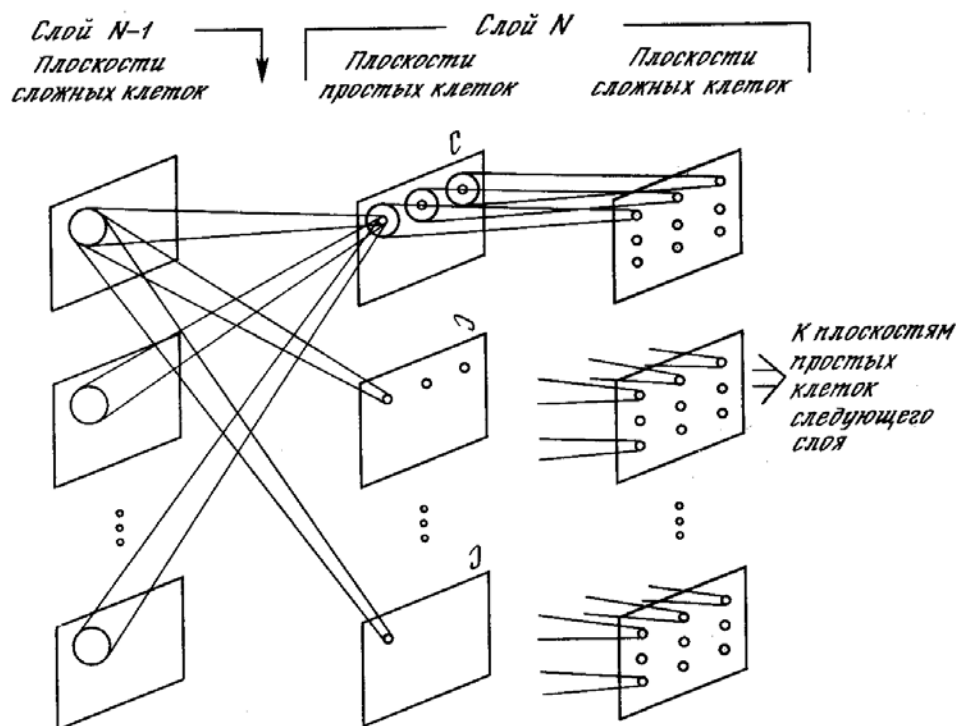


Рис. 3.23. Неокогнитрон

Каждый слой состоит из плоскостей простых и сложных клеток. Каждый нейрон простой плоскости связан с локальным

двумерным участком плоскостей предыдущего слоя, веса всех нейронов в пределах одной плоскости одинаковы, и таким образом плоскость реагирует на определённый образ, находящийся в участке изображения (пример на рисунке – плоскости реагируют на букву «С» повернутую под разными углами). Положение активированного таким образом нейрона в простой плоскости отмечает участок, в котором найден этот образ, независимо от искажения этого образа. Нейрон сложной плоскости связан с участком своей простой плоскости, и обнаруживает активность нейронов на этом участке, уменьшая таким образом чувствительность к позиции образа.

Таким образом достигается иерархическая обработка изображения, когда на последующих слоях неокогнитрон реагирует на более общие черты изображения не сбиваясь на искажения, сдвиг и т.д.

Классический неокогнитрон является мощным средством распознавания изображений, однако требует высоких вычислительных затрат, которые на сегодняшний день недостижимы [42].

Однако существует множество работ, например [12, 20, 89], направленных на совершенствование неокогнитрона. Один из наиболее перспективных подходов для распознавания человека по изображению лица – это свёрточные нейронные сети [12,20].

3.4.3. Свёрточные нейронные сети

В классической многослойной нейронной сети межслойные нейронные соединения полносвязны, и изображение представлено в виде n -мерного вектора, не учитывающего ни двумерную локальную организацию пикселей, ни возможностей деформации. Архитектура свёрточной НС (рис. 3.24) направлена на преодоление этих недостатков и основывается на принципах архитектуры неокогнитрона, упрощённого и дополненного обучением алгоритмом обратного распространения ошибки [20].

В ней использовались локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов), общие веса (обеспечивают детектирование некоторых черт в любом месте изображения) и иерархическая организация с пространственными подвыборками (*spatial subsampling*).

Свёрточная НС (СНС, Convolutional Neural Network) обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

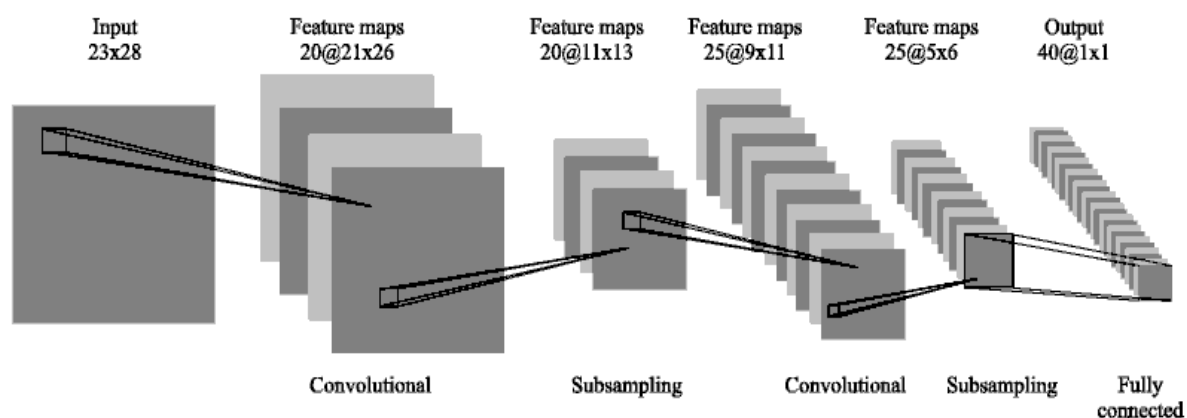


Рис. 3.24. Архитектура свёрточной нейронной сети

Архитектура СНС, (рис. 3.24), состоит из многих слоёв. Слои бывают двух типов: свёрточные (*Convolutional*) и подвыборочные (*Subsampling*), свёрточные и подвыборочные слои чередуются друг с другом. В каждом слое имеется набор из нескольких плоскостей, причём нейроны одной плоскости имеют одинаковые веса, ведущие ко всем локальным участкам предыдущего слоя (как в зрительной коре человека), изображение предыдущего слоя как бы сканируется небольшим окном и пропускается сквозь набор весов, а результат отображается на соответствующий нейрон текущего слоя. Таким образом набор плоскостей представляет собой карты характеристик (*feature maps*) и каждая плоскость находит «свои» участки изображения в любом месте предыдущего слоя. Следующий за свёрточным слоем подвыборочный слой уменьшает масштаб плоскостей путём локального усреднения значений выходов нейронов. Таким образом, достигается иерархическая организация. Последующие слои извлекают более общие характеристики, меньше зависящие от искажений изображения. Обучается СНС стандартным методом обратного распространения ошибки. Сравнение МНС и СНС [20] показало существенные преимущества последней как по скорости, так и по надёжности классификации. Полезным свойством СНС является и то, что характеристики, формируемые на выходах верхних слоёв иерархии, могут быть применимы для классификации по методу ближайшего соседа (например, вычисляя Евклидово расстояние), причём СНС может успешно извлекать такие характеристики и для образов, отсутствующих в обучающем наборе. Для СНС характерны высокая скорость обучения и работы. Тестирование СНС на базе данных ORL, содержащей изображения лиц с небольшими изменениями освещения, масштаба, пространственных поворотов, положения и различными эмоциями, показало 98%-ную точность распознавания.

3.5. Достоинства и недостатки нейросетевых методов распознавания

Рассмотренные нейросетевые методы обеспечивают быстрое и надёжное распознавание изображений. Однако при применении этих методов к изображениям трёхмерных объектов возникают трудности, связанные с пространственными поворотами и изменением условий освещённости.

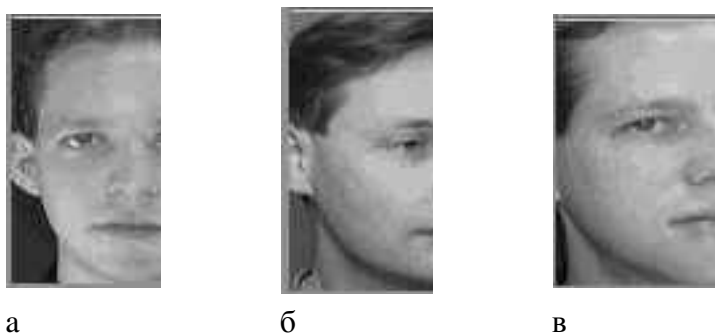


Рис. 3.25. С точки зрения классифицирующей системы (в) больше похоже на (б) чем на (а)



Рис. 3.26. Изображения одного и того же лица существенно различается при изменении освещения.

Изображения для различных углов поворота объекта существенно различаются, часть информации на изображении теряется, возникает новая, специфическая для данного угла [10,31]. Например, лицо, повернутое на некоторый угол, с точки зрения классифицирующей системы больше похоже на лицо другого человека, повернутое на такой же угол, чем на нужное лицо, изображённое в фас (рис. 3.25.). Аналогичная проблема с изменением освещения (рис. 3.26.). Такие ограничения обычно преодолеваются путём предъявления всевозможных вариаций изображения (различные повороты и освещённость) при обучении, но построение такого обучающего набора - трудная задача, и чаще всего такие наборы недоступны. Как показывает мировой опыт, эти проблемы не могут быть полностью решены выбором исходного представления данных. Поэтому к классифицирующим системам предъявляется требование - имея конечный репрезентативный набор вариаций образов некоторых классов, обобщить свой опыт на все остальные классы, не входившие в обучающий набор. Т.е. система должна

извлечь характеристики, инвариантные к внутриклассовым изменениям и максимально репрезентативные по отношению к межклассовым изменениям. Такая задача в общем виде для систем распознавания лиц ещё не решена, но существуют методы, которые показывают возможности решения отдельных её аспектов (инвариантность к освещению, синтез повёрнутых в пространстве изображений лиц на основе обучения).

Так же существуют трудности, связанные с внутриклассовыми вариациями. Для лиц это различные эмоции, закрытые/открытые глаза, наличие очков и бород, изменения в причёске. Эти случаи система так же должна уметь обобщать.

В общем случае, при распознавании человек использует информацию от различных источников, и кроме того привлекает огромный запас контекстных знаний, который системам распознавания образов пока недоступен.

3.6. Применение ИНС для извлечения ключевых характеристик

3.6.1. Применение ИНС для извлечения ключевых характеристик лица

Постановка задачи распознавания лица. Входным классом для алгоритма распознавания был класс нормализованных лиц, то есть подразумевалось то, что в изображении выделена область лица, выполнена нормализация освещенности, определен ракурс изображения лица и произведено согласование размеров изображений. Задача биометрической системы идентификации состоит в том, чтобы принять правильное решение в пользу одной из двух гипотез:

- H_0 - зарегистрированное лицо, которое система должна принять;
- H_1 - незарегистрированное лицо, которое система должна отвергнуть.

Тогда ошибка первого рода, при которой мы отвергаем гипотезу H_0 и принимаем H_1 , будет характеризовать вероятность ошибочных отказов – FRR (false rejection rate), то есть процент зарегистрированных пользователей, отвергнутых системой. Ошибка второго рода, при которой мы принимаем H_0 и отвергаем H_1 будет характеризовать вероятность ошибочных подтверждений – FAR (false acceptance rate), то есть процент незарегистрированных пользователей, пропущенных системой [12].

Метод распознавания изображения включает следующие

процедуры:

- преобразование исходного изображения в начальное представление алгоритма (предобработка, сегментация и математические преобразования)
- механизм обобщения, сравнения, классификации признаков изображения, и принятия решений: кластерная модель, метрика, нейронная сеть.

При этом для правильного распознавания изображение должно пройти несколько этапов предобработки для устранения всех неинформативных компонент:

Поиск лица на изображении;

Поворот изображения (центры зрачков должны находится на одной прямой)

Масштабирование (нормализация по расстоянию между зрачками)

Кадрирование(вырезание прямоугольной области с заданными размерами содержащей лицо)

Выравнивание яркостных характеристик

Выделение информации, требуемой для дальнейшего процесса распознавания

На рис. 3.27 приведена схема системы распознавания личности человека по изображению его лица [48].

3.7. Применение ИНС для классификации образов

Системы распознавания образов на основе ИНС делятся 2 основных класса: системы с распознаванием напрямую по входным сигналам и системы с заранее выделенными ключевыми характеристиками.

В ИНС для классификации напрямую по входным сигналам ключевые характеристики извлекаются в неявном виде внутри самой ИНС, что существенно осложняет процесс понимания функционирования подобных систем.

Поэтому наибольшее распространение получили системы второго типа

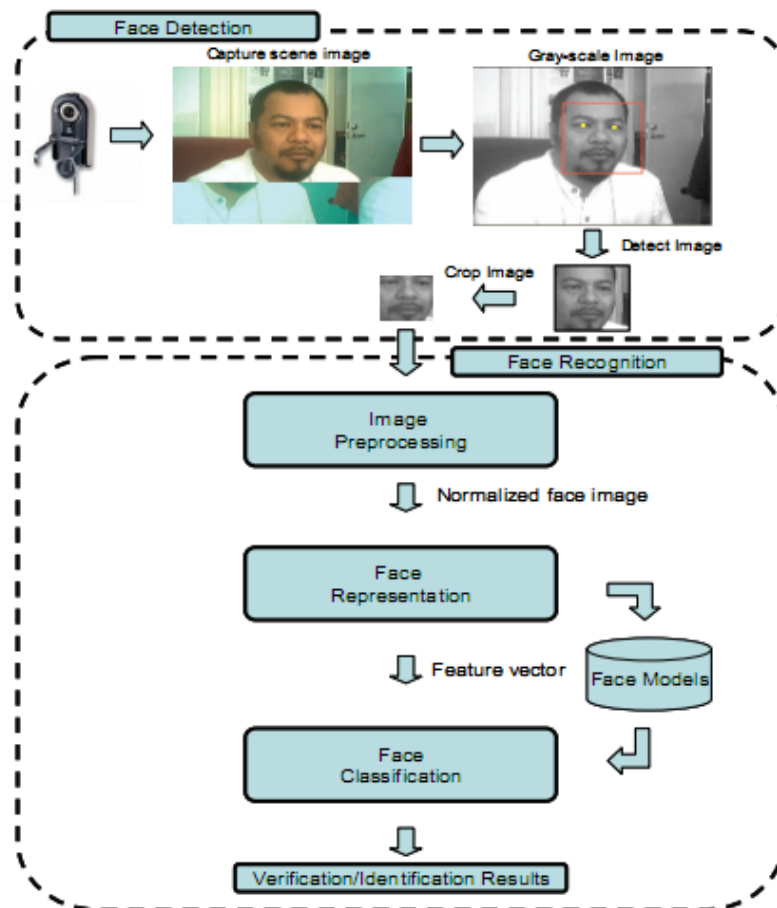


Рис. 3.27. Схема системы распознавания личности по изображению лица

3.7.1. . Применение ИНС для классификации напрямую по входным сигналам

В работах [20,48] рассматриваются варианты построения систем распознавания лиц следующей структуры (рис. 3.28.):

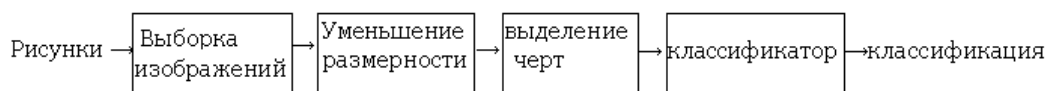


Рис 3.28. Блок-схема системы распознавания лиц

Исследовалось влияние каждого блока на точность процесса распознавания лиц.

В блоке выборки изображения провели оценку двух разных методов представления локальных шаблонов изображений. В каждом методе окно перемещается по изображению, как показано на рис. 3.29.

Первый метод просто создает вектор данных из набора в локальном окне изображения используя значения интенсивности в каждой точке окна. Пусть $X_{i,j}$ это будет интенсивность точки в i -й

колонке и j -й строке данного изображения. Если локальное окно квадратное или стороны длины $2W+1$, где центром является $X_{i,j}$ тогда вектор это просто

$$[X_{i-W,j-W}, X_{i-W,j-W+1}, \dots, X_{i,j}, \dots, X_{i+W,j-W-1}, X_{i+W,j+W}].$$

Второй метод создает представление локального шаблона формируя выходной вектор из интенсивности центрального пикселя $X_{i,j}$ и разнице в интенсивности между центральным пикселем и всеми другими пикселями в квадратном окне. Вектор задается как

$[X_{i,j}-X_{i-W,j-W}, X_{i,j}-X_{i-W,j-W+1}, \dots, wijX_{i,j}, \dots, X_{i,j}-X_{i+W,j-W-1}, X_{i,j}-X_{i+W,j+W}]$. Результирующее представление становится частично нечувствительным к вариациям в интенсивности всего шаблона. Степень инвариантности может быть модифицирована установлением веса wij привязанному центральному компоненту интенсивности.

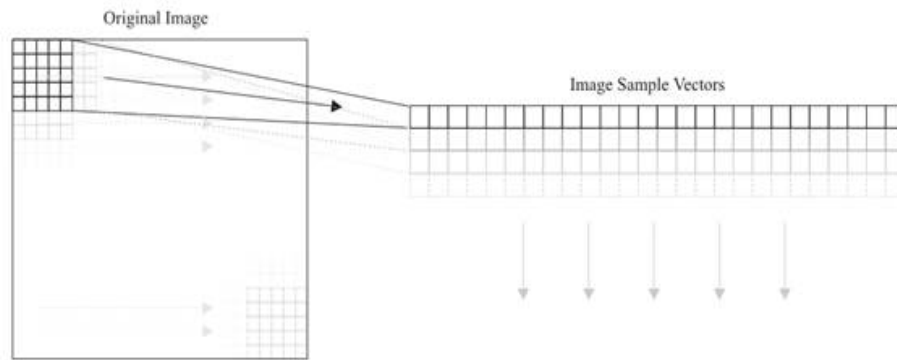


Рис. 3.29. – Процесс создания локального шаблона. Окно делает шаги по рисунку и создаются вектора для каждого положения окна.

Для уменьшения размерности использовались следующие алгоритмы.

Самоорганизующаяся карта Кохонена (SOM): алгоритм определяет карты из входного множества R_n на топологически определенный набор узлов, обычно в пространстве, имеющем меньше измерений. Каждый из узлов описывается двумя векторами, первый — вектор веса m_i , имеющий такую же размерность, что и входные данные. Второй — координаты узла на карте. Пример двумерного SOM показан на рис. 3.30. Вектор-ссылка во входном множестве,

$$\mathbf{m}_i \equiv [\mu_{i1}, \mu_{i2}, \dots, \mu_{in}]^T \in \mathcal{R}^n, \quad (3.13)$$

привязан к каждому узлу в SOM. Во время обучения каждый входной вектор \mathbf{x} сравнивается со всеми m_i в поиске размещения наиболее сходного m_c которое задается следующим образом

$$|\mathbf{x} - \mathbf{m}_c| = \min_i \{|\mathbf{x} - \mathbf{m}_i|\}.$$

Узлы в SOM пересчитываются согласно:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \quad (3.14)$$

где t это время, в течении которого уже происходит обучение, $h_{ci}(t)$ – «ближайшая функция» - сглаживающая функция, максимум которой находится в m_c . Обычно

$$h_{ci}(t) = h(|\mathbf{r}_c - \mathbf{r}_i|, t)$$

где \mathbf{r}_c и \mathbf{r}_i представляют положение узлов в выходном пространстве SOM. \mathbf{r}_c – это узел с наиболее близким весовым вектором к входному шаблону, а \mathbf{r}_i – пробегает по всем узлам. $h_{ci}(t)$ приближается к 0, когда $|\mathbf{r}_c - \mathbf{r}_i|$ возрастает и также t стремится к бесконечности. Примененная близкая(аппроксимирующая?) функция это:

$$h_{ci} = \alpha(t) \exp\left(-\frac{|\mathbf{r}_c - \mathbf{r}_i|^2}{2\sigma^2(t)}\right) \quad (3.15)$$

где $\alpha(t)$ - это коэффициент обучаемости и $\sigma(t)$ определяет длину ряда. Они обычно оба монотонно убывают с течением времени. Использование аппроксимирующей функции означает, что узлы, которые располагаются в SOM структуре в соответствии с положением «победившего» узла. Это создаёт сглаживающий эффект, который приводит к глобальной организации карты. SOM может быть представлена, как нелинейная проекция плотности вероятностей [58,71].

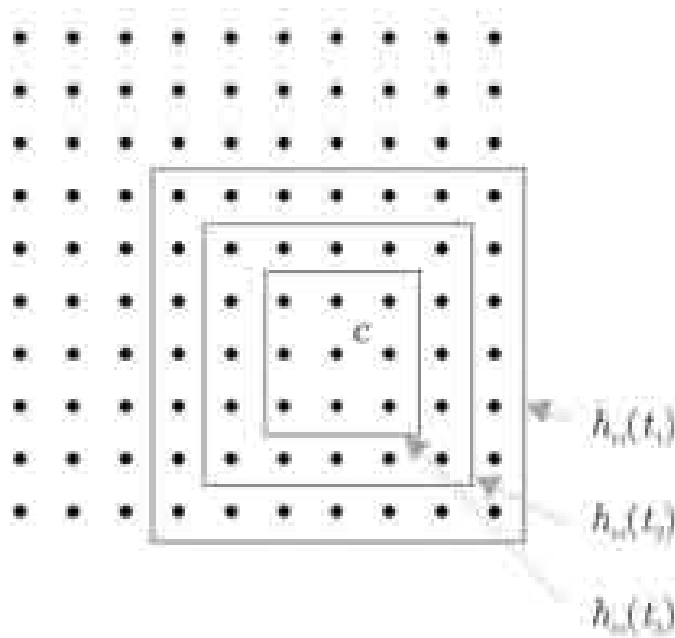


Рис 3.30. — двумерная SOM, показывающая квадратичную аппроксимирующую функцию с начальным значением $h_{ci}(t_1)$ и сокращающаяся до размера $h_{ci}(t_3)$ со временем.

Второй метод использованный для уменьшения размерности входных данных - преобразования Кархунена-Лоева (ПКЛ): оптимальный линейный метод для избежания избыточности исходных данных — преобразование Кархунена-Лоева (ПКЛ) или расширение собственного вектора с помощью Метода Главных Компонент (МГК). МГК генерирует набор ортогональных осей проекций, называемых главными компонентами или собственными векторами в распределении входящих данных для сокращения разброса данных. ПКЛ является хорошо известным статистическим методом для извлечения черт и многомерных проекций данных и был широко используем в распознавании образов, обработки сигналов, изображений, анализа информации. Точки в n -мерном входном пространстве проецируются на m -мерное пространство, $m \leq n$. ПКЛ используется в нашей работе для сравнения с SOM. В уменьшении размерности локальных наблюдений изображений. ПКЛ также используется в технологии Eigen Faces, но, в этом случае, оно использовано для всего изображения, в то время, как мы используем его для малых локальных наблюдений изображения.

Для выделения основных черт и классификации изображений применялись сверточные нейронные сети. Проблема распознавания лиц на двумерных изображениях обычно очень некорректно, так как существуют много моделей, которые могут подходить к обучающим точкам, но не проводят обобщение для новых изображений. Другими словами, недостаточно обучающих точек, созданных в пространстве входящими изображениями для того, чтобы гарантировать точное разделение классов на входящем множестве. В дополнение для многослойного персептрона, принимающего двумерные изображения, как входящие, не существуют чувствительности к локальным деформациям.

Сверточные нейросети включают ограничения и достигают некоторой степени инвариантности к деформациям, используя три основные идеи: локально восприимчивые поля, разделённые веса и пространственные подвыборки. Использование разделённых весов также сокращает число параметров в системе поддержки обобщения. Сверточные нейросети успешно применялись для распознавания символов. Сеть состоит из набора слоёв, которые содержат один или более матриц. Примерно отцентрированные и нормированные изображения поступают во входной слой. Каждый элемент матрицы получает на вход небольшое окружение (наблюдение) элемента матрицы из предыдущего слоя. Идея привязки элементов к локально восприимчивым полям датируется 60ми годами — временем развития идеи персептрона и открытия локальной чувствительности (например, ориентационно - избирательных нейронов в зрительной системе кошки). Веса, формирующие восприимчивое поле для матрицы

предполагаются равными во всех точках матрицы. Каждая матрица может быть представлена, как карта черт, которая имеет фиксированный детектор черт, который сворачивает локальное окно, которое было снято с образа на предыдущем слое. Много образов-матриц используется на каждом слое для распознавания большого числа черт. Эти слои называются свёртывающими слоями. Как только черта была определена, её точное положение становится менее важным. Таким образом, сворачивающие слои обычно дополняются следующими за ними слоями, которые выполняют распределяющие и отбрасывающие операции. Сеть обучается с помощью обычного обучения с обратной связью без учителя с использованием градиентного спуска. Стратегия связи может быть использована для уменьшения числа весов в сети [12,20].

Сама система распознавания личности человека по его портрету имеет архитектуру показанную на рис. 3.31 [22]:

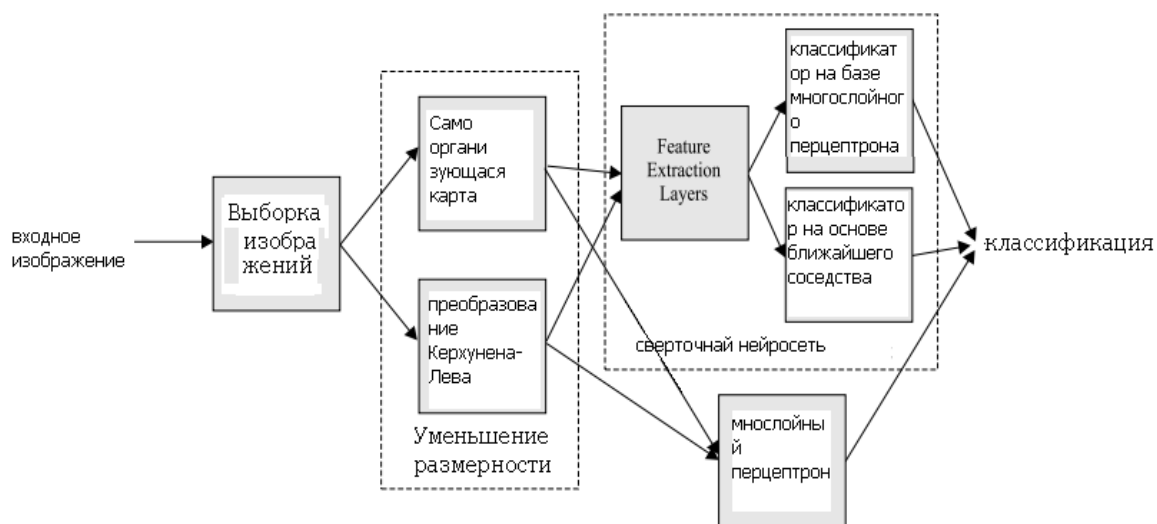


Рис. 3.31. Архитектура системы распознавания личности по портрету.

Система работает следующим образом.

Для изображений в обучающем наборе окно фиксированного размера (5x5) передвигается по исходному изображению, как показано на рис. () и локальные наблюдения извлекаются на каждом шаге. Шаг составляет 4 пикселя.

Самоорганизующаяся карта (с тремя измерениями и пятью узлами для каждого измерения, $5^3 = 125$ узлов) обучается на векторах, полученных на предыдущем шаге. SOM квантует 25-мерные входящие вектора в 125 топологически обусловленных значений. Три измерения SOM могут рассматриваться, как три черты. Мы также экспериментируем с заменой SOM преобразованием Кархунена-Лова. В этом случае, ПКЛ преобразует вектора в 25-мерном

пространстве в 3-х мерное пространство.

То же окно, что и на первом шаге проходит все изображения в обучающем и тестовом множествах. Локальные видения изображения передаются через SOM на каждом шаге, формирую новые обучающие и тестовые группы в выходном множестве, созданном самоорганизующейся картой. Каждое входящее изображение теперь представляется тремя картами, каждое из которых отвечает измерению SOM. Размеры этих карт равны размерам входного изображения — 92x112 разделённого размером шага.

Сверточная ИНС обучается на созданном обучающем множестве. Мы также экспериментировали с обучением стандартного многослойного персептрона для сравнения. структура сверточной нейронной сети представлена на рис. 3.32.

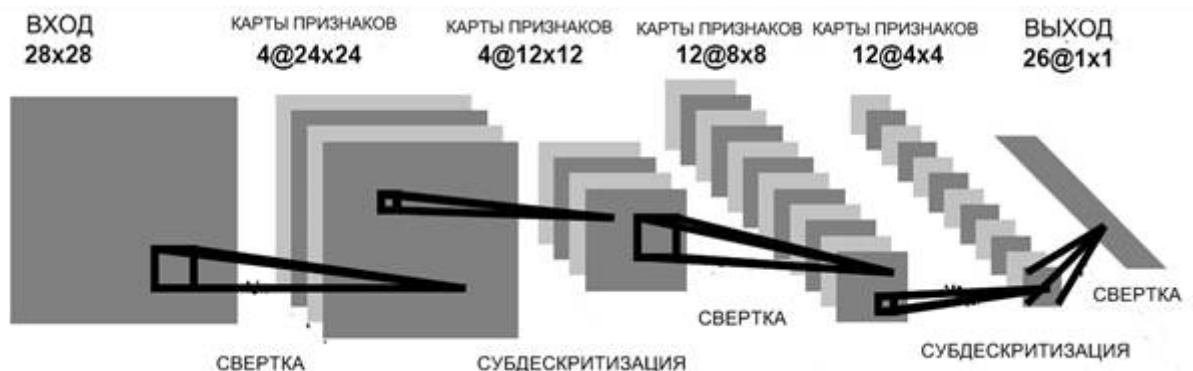


Рис. 3.32. Сверточная нейронная сеть

Сверточная нейронная сеть состоит из чередующихся слоев свертки и субдискретизации. Поступающее на вход изображение подвергается свертке с некоторым ядром свертки в соответствии с выражением:

$$C\{n\}(i, j) = F \left[\sum_{k=1}^{k_s} \sum_{l=1}^{l_s} K\{n\}(k, l) \cdot S(i-k, j-l) + B\{n\} \right] \quad (3.16)$$

В формуле (3.16) C – результат свертки (карта признаков); n – номер слоя; i, j–индексы, определяющие положение элемента в карте признаков; F – функция насыщения, обычно задаваемая как сигмоидальная или тангенциальная функция; k, l – индексы определяющие положение элемента в матрице ядра свертки; ks, ls – размеры ядра свертки; K– ядро свертки; S – входное изображение; B – матрица смещений.

Ядро свертки является набором разделяемых весовых коэффициентов. Результатом данной операции является также некоторое изображение, которое называется картой признаков. В зависимости от выбранного ядра свёртки, карта признаков будет выделять те или иные характеристики входного изображения. Для

наиболее полного выделения характеристик входного изображения используется несколько различных ядер свертки так, что на выходе сверточного слоя получается несколько карт признаков.

За сверточным слоем следует слой усреднения и субдискретизации, который понижает размерность карты признаков, тем самым понижая чувствительность выходов к сдвигам и поворотам [20].

Сверточная нейросеть состоит из пяти слоёв, не включая входной. Коэффициент уверенности вычисляется для каждой классификации $u_m(u_m - u_{2m})$, где u_m — максимальный выход, u_{2m} — второй максимальный выход. Число матриц в каждом слое, мерность матриц, мерность восприимчивых полей показана в таблице 3.1:

Таблица 3.1.

Измерения для сворачивающей нейросети.

Слой	Тип	Элементы	X	Y	Восприимчивое поле x	Восприимчивое поле y	Процент связей*
1	Свёртывающий	20	21	26	3	3	100
2	Связывающий	20	11	13	2	2	-
3	Свёртывающий	25	9	11	3	3	30
4	Связывающий	25	5	6	2	2	-
5	Полностью связанный	40	1	1	5	6	100

*) Процент связей — процент узлов в предыдущем слое, с которым каждый нейрон в текущем слое соединён.

Сеть обучалась с помощью алгоритма обратной связи (20000 шагов). Веса в сети обновлялись после каждого представления видения. Все выходные значения были нормализованы (от -1 до 1). Лучшие из 10 случайных наборов весов выбирались для начальной инициализации параметров сети. Веса были инициализированы в узле случайным числом в пределах от $-2.4/F_i$, $2.4/F_i$, F_i — коэффициент объединения по входу нейрона i .

$$\eta = \frac{\eta_0}{\frac{n}{N/2} + \frac{c_1}{\max\left(1, \left(c_1 - \frac{\max(0, c_1(n - c_2 N))}{(1 - c_2)N}\right)\right)}} \quad (3.17)$$

где N — Общее количество эпох обучения, n — текущая эпоха, $c_1 = 50$, $c_2 = 0.65$

Сравнение различных архитектур системы распознавания лиц с другими вариантами показано в таблице 3.2 [90].

Таблица 3.2.

Ошибка распознавания различных архитектур системы

Номер варианта	Вариант архитектуры системы	Средняя ошибка распознавания
1	мерность карты Кохонена 1\2\3\4	8.25\6.75\5.75\5.83%
2	число узлов на измерение SOM 4\6\8\10	8.5\6\3.83\ 4.16%
3	изменение алгоритма выборки изображений (алгоритм 1\алгоритм 2)	5.75\7.17%
4	замена карты SOM алгоритмом Кархунена-Лоэва	5.33
5	оптимальная структура системы (3-мерная SOM, 8 узлов на измерение, выборка согласно алгоритму 1, сверточная сеть для распознавания)	3.83%
6	замена сверточной сети многослойным перцептроном	5.85%

Результаты работы системы распознавания в базе лиц ORL показаны на рис.3.33.



Рис. 3.33. Тестовые изображения. Изображения в белой рамке – неправильно классифицированные

Сравнение характеристик системы распознавания оптимальной архитектуры с известными системами распознавания лиц показано в таблице 3.3. [20]

Таблица 3.4. показывает результаты при изменении числа изображений для каждого класса, использованных в обучающем множестве (от 1 до 5) для разных алгоритмов [20].

Таблица 3.3.

Сравнение систем распознавания лица

Система	Ошибка распознавания	Время распознавания
Собственных лиц (eigenfaces)	10.5%	неизвестно
Псевдо-двумерная скрытая марковская модель	5%	240 сек на Sun Sparc II
Замена сверточной сети многослойным перцептроном	3.83%	<0.5 сек на SGI Indy MIPS R4400 100Mhz

Таблица 3.4.

Зависимость ошибки распознавания от числа изображений в классе

Images per person	1	2	3	4	5
Eigenfaces – average per class	38.6	28.8	28.9	27.1	26
Eigenfaces – one per image	38.6	20.9	18.2	15.4	10.5
PCA+CN	34.2	17.2	13.2	12.1	7.5
SOM+CN	30.0	17.0	11.8	7.1	3.8

В работе [22] для распознавания лиц использовалась непосредственно сверточная нейронная сеть, на вход которой подается фотопортрет человека (рис. 3.34.). Далее происходит сканирование входящего изображения локальным рецептивным полем, информация каждого рецептивного поля преобразуется функцией и записывается в каждый нейрон следующего слоя. Этот процесс продолжается в каждом слое нейронной сети. Таким образом, нейронная сеть извлекает характеристики изображения, инвариантные к масштабу изображения, ракурсу фотоснимка, повороту и т.д. На выходе устанавливается соответствие одному из классов изображений, хранимых в базе данных.

Точность распознавания подобной системой достигала 97.5 %, при достаточной устойчивости к зашумлению входного сигнала.

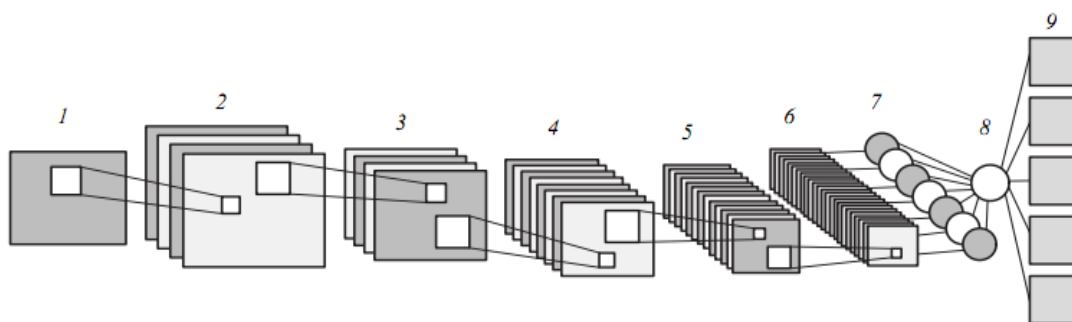


Рис. 3.34. Архитектура сверточной нейронной сети для распознавания лиц. 1 – вход, 2,4,6 – сверточные слои, 3,5 – предвыборочные слои, 7 – слой обычных нейронов, 8 – выходной слой, 9 – классы изображений лиц

3.7.2. Системы распознавания лиц напрямую по входным сигналам с неявным извлечением ключевых характеристик внутри сети

В данном случае речевой сигнал является входными данными нейронной сети. После обработки звуковых данных получен массив сегментов сигналов. Каждый сегмент соответствует набору чисел, характеризующих амплитудные спектры сигнала. Для подготовки к вычислению для сигнала выхода нейронной сети необходимо записать все наборы чисел в таблицу, строка которой – это набор чисел каждого кадра (табл. 3.5) [21]. В указанной таблице I – количество значений одного набора чисел, N – количество наборов чисел (кадров сигнала после нарезки)

Количество входных и выходных нейронов известно. Каждый из входных нейронов соответствует одному набору чисел. В выходном слое только один нейрон, выход которого соответствует желаемому значению распознавания сигнала.

Таблица 3.5.

Представление речевого сигнала набором чисел.

Кадр	1-ое значение	2-ое значение	...	I -ое значение
1-ый кадр	x_{11}	x_{12}	...	x_{1I}
2-ый кадр	x_{21}	x_{22}	...	x_{2I}
...
N -ый кадр	x_{N1}	x_{N2}	...	x_{NI}

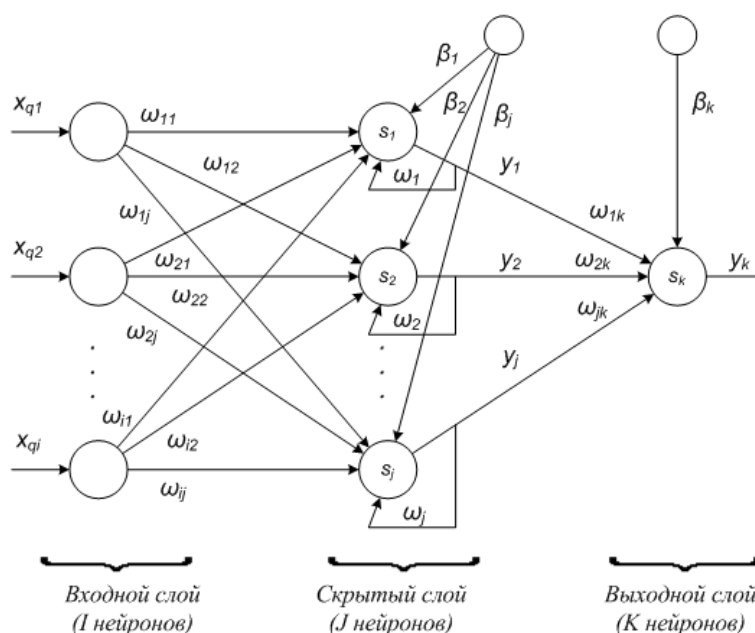


Рис. 3.35. – Структура нейронной сети с одной обратной связью

На рис.3.35 x_{qi} – i -ое входное значение q -го набора чисел;

y_j – выход j -го нейрона слоя; w_{ij} – весовой коэффициент связи, соединяющей i -ый нейрон с j -ым нейроном; w_j – весовой коэффициент обратной связи j -го нейрона; b_j – смещение j -го нейрона слоя.

Для вычисления выхода нейронной сети необходимо выполнить следующие последовательные шаги:

Шаг 1: Инициировать все контексты всех нейронов скрытого слоя .

Шаг 2: Подать первый набор чисел на вход нейронной сети . Вычислить для него выходы скрытого слоя.

$$y_j = f\left(\sum_{i=1}^I \omega_{ij}x_{1i} + \beta_j + \omega_j x_j\right) \quad (3.18.)$$

где $f(x)$ – нелинейная активационная функция, вследствие чего можно записать

$$y_j = \frac{1}{1 + e^{-\alpha S_j}}$$

Шаг 3: Если текущий набор чисел не является последним, то переход на шаг 5, иначе переход на шаг 4.

Шаг 4: Записать выходы нейронов скрытого слоя на контексты $x_j = y_j$, где . Переход к шагу 2 для следующего набора чисел.

Шаг 5: Вычислить выход нейрона выходного слоя.

$$y_k = f\left(\sum_{j=1}^J \omega_{jk}y_j + \beta_k\right) \quad (3.19.)$$

Рассмотрим задачу, которая состоит в распознавании чисел от 0 до 9. Для распознавания одного числа нужно построить собственную нейронную сеть. И так должно построить 10 нейронных сетей. Надиктована база из 250 слов (числа от 0 до 9) с различными вариациями произношения. База случайным образом разделялась на две равные части – обучающую и тестирующую выборки. При обучении нейронной сети распознаванию одного числа, например 5, желаемый выход этой нейронной сети должен быть единицей для обучающей выборки с числом 5, а остальные – нулю.

Обучение нейронной сети осуществляется путем последовательного предъявления обучающей выборки, с одновременной подстройкой весов в соответствии с определенной процедурой, пока ошибка настройки по всему множеству не достигнет приемлемого низкого уровня. Функции ошибки в системе будет вычисляться по следующей формуле:

$$E = \frac{1}{2N} \sum_{i=1}^N (y_{ki} - d_i)^2 \quad (3.20.)$$

где N – количество обучающих выборок, обработанных нейронной сетью примеров;

- реальный выход нейронной сети;
- желаемый (идеальный) выход нейронной сети.

Для каждого слова из тестовой выборки реальные выходы вычисляются 10 нейронными сетями распознавания разных чисел. Нейронная сеть, которая имеет максимальное выходное значение, и является нейронной сетью распознавания данного слова. И слово, распознанное нейронной сетью, является результатом распознавания.

В работе [12] рассматривались различные архитектуры ИНС для системы распознавания речи – использовался многослойный персептрон с различным количеством скрытых слоев (рис.3.36). Трехслойный персептрон показал точность на 1-3% выше остальных архитектур.

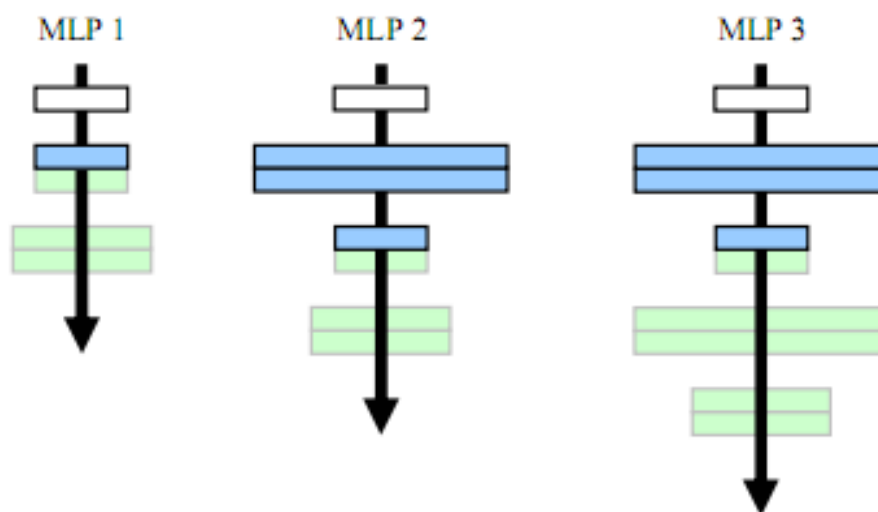


Рис. 3.36 Архитектуры, использовавшиеся для распознавания речи.
(отмеченные светло-голубым слои использовались только для обучения)

В работе [36] рассмотрена робастность (устойчивость) системы распознавания с неявным внутрисетевым извлечением характеристик к шумам. ИНС на основе многослойного персептрона использовалась для распознавания речи в условиях наложения шумов различной природы и интенсивности. В том случае если в обучающей выборке оказывался зашумленный сигнал правильность распознавания значительно повышалась .

3.7.3. Применение ИНС для классификации по заранее извлеченным признакам

В данных системах сначала используются различные алгоритмы для выделения ключевых характеристик, после чего они поступают на вход ИНС для дальнейшей классификации

3.7.3.1. Метод собственных фильтров для распознавания лица

Построение фильтров методом главных компонент указано на рис. 3.37.



Рис. 3.37 Собственные фильтры, полученные в результате самоорганизации на БД изображений лиц и результат выделения целостных признаков

Для определения целостных признаков на изображении используется модифицированный метод главных компонент. На изображении лица случайным образом выбирались блоки размером 16×16 . По собранной статистике для 10000 блоков строились собственные фильтры, использовавшиеся потом для извлечения черт на изображении. Рассмотрим метод более детально. Пусть x - вектор размерности $N \times 1$, соответствующий изображению $N=16 \times 16$ и задана статистика векторов x (200 изображений размером $16 \times 16=256 \times 1$)

u – вектор той же размерности (256×1) что и x , на него мы будем проектировать входные вектора x :

Пусть

$$\begin{aligned}\bar{x} &= E[\{x\}] \\ a &= (x - \bar{x})^T u = u^T (x - \bar{x})\end{aligned}\quad (3.21)$$

– проекция X на U.

$$\sigma^2 = E[a^2] = E[(u^T (x - \bar{x}))((x - \bar{x})^T u)] = u^T E[(x - \bar{x})(x - \bar{x})^T] u = u^T R u \quad (3.22)$$

дисперсия для данной статистики образов и есть функция от u:

$$\psi(u) = \sigma^2 = u^T R u,$$

где $R = E[(x - \bar{x})(x - \bar{x})^T]$ – матрица ковариации размерности $N \times N$.

Задача анализа главных компонент выбрать базис пространства $u = (u_1, u_2, \dots, u_k)$ так, чтобы вариация (дисперсия) проекции была минимальной. Минимум этой функции достигается решением задачи на собственные значения $Ru = \lambda u$. На Рис.4.37 приведены собственные фильтры, полученные в результате самоорганизации на БД изображений лиц и результат выделения целостных признаков, показан результат применения метода главных компонент к блокам изображений результат их самоорганизации в определенные типы фильтров, а также результат фильтрации – первые 10 разных компонент.[74,86,101]

Основное преимущество применения анализа главных компонент – это хранение и поиск изображений в больших базах данных, реконструкция изображений.

Основной недостаток – высокие требования к условиям съёмки изображений. Изображения должны быть получены в близких условиях освещённости, одинаковом ракурсе и должна быть проведена качественная предварительная обработка, приводящая изображения к стандартным условиям (масштаб, поворот, центрирование, выравнивание яркости, отсечение фона). Нежелательно наличие таких факторов, как очки, изменения в причёске, выражении лица и прочих внутриклассовых вариаций

3.7.3.2. Метод сравнения эластичных графов на основе вейвлетов Габора.

Для определения локальных признаков на изображении использовались фильтры на основе вейвлетов Габора. Разложение Габора было предложено в качестве инструмента анализа текстур. Этот подход был мотивирован современными представлениями о психофизической природе человеческого зрения. Декомпозиция Габора представляет собой так называемый «Направленный микроскоп» с набором ориентаций и масштабам.

В результате преобразования Габора спектральная плоскость

разбивается на m частотных и n ориентационных диапазонов [79,104,105]:

$$\hat{g}_{ij}(\omega) = \exp(-1/2(\omega - \omega_{ij})^t R_j \Sigma_{ij}^{-1} (\omega - \omega_{ij})) \dots$$

- вейвлет, соответствующий (i,j) ориентации и масштабу, где $1 \leq i \leq m$, $1 \leq j \leq n$, R_j – матрица поворота, а Σ_{ij} – диагональная матрица.

$A_{ij} = R \Sigma_{ij}^{-1} R^t$ - положительно определенная матрица, главные оси которой определяют эллипс, ориентированный в соответствии с направлением, задаваемым j . Конкретнее,

$$R_j = \begin{pmatrix} \cos(\phi_j) & \sin(\phi_j) \\ -\sin(\phi_j) & \cos(\phi_j) \end{pmatrix}$$

$$\Sigma_{ij} = \begin{pmatrix} \sigma_{r_i}^2 & 0 \\ 0 & \sigma_{\phi_j}^2 \end{pmatrix}$$

$$\omega_{ij} = R_j \begin{pmatrix} \omega_{r_i} \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\phi_j) \omega_{r_i} \\ \sin(\phi_j) \omega_{r_i} \end{pmatrix} \quad (3.23)$$

где ω_{ij} - центральная частота фильтра.

Благодаря разложению Габора, вычисление соответствующей функции точечного распределения можно провести аналитически. Кроме описанного разложения можно также использовать другие разложения с использованием с использованием логарифмическо - полярных отображений. Точечный отклик, соответствующий \hat{g}_{ij} , вычисляется по формуле:

$$g_{ij}(x) = \exp(-1/2x^t R_j \Sigma_{ij} R_j^t x) \exp(i2\pi \omega_{ij}^t x) \quad (3.24)$$

Займемся определением остальных параметров. Разложение по n направлениям и m масштабам выбирается так, чтобы соседние фильтры пересекались под одинаковыми значениями $1/\sqrt{e}$ вдоль своих главных осей. Направления выбирают эквидистантными, вследствие чего значения σ_{ϕ_j} постоянны для всех j :

$$\sigma_{\phi_j} = \sigma_{\phi} = \pi/2n$$

Угол между фильтрами равен π/n , направления задаются формулой:

$$\omega_{\phi_j} = 2\sigma_{\phi}(j-1)$$

Реальные диапазоны частот распределены так, что частотный диапазон удваивается на каждом шаге, покрывая интервал от $\omega_{rmin} > 0$ до $\omega_{rmax} < 1/2$ в нормализованных частотах, где единица соответствует частоте Найквиста [88].

Определим промежуточную переменную:

$$\sigma_0 = (\omega_{rmax} - \omega_{rmin}) / (2(2^m - 1))$$

Тогда радиальные центры задаются формулами:

$$\omega_{rj} = \omega_{rmin} + \sigma_0(1 + 3(2^{j-1} - 1)), \quad \sigma_{rj} = \sigma_0 2^{j-1} \quad (3.25)$$

Параметры максимальной и минимальной центральной частоты вейвлетов Габора: $\omega_{rmin} = 1/16$, $\omega_{rmax} = 1/4$

Ориентация вейвлетов выбиралась эквидистантно с круговой

частотой – для 6 ориентаций, то есть центры эллипсов, характеризующие частотную область действия вейвлета, задавались со следующими ориентациями[103,104]:

$$\omega_{\phi_j} = 2x\sigma_{\phi}(j-1) = \{0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6\}$$

На рис. 3.38 показаны фильтры Габора и результат их применения локальные признаки изображения.

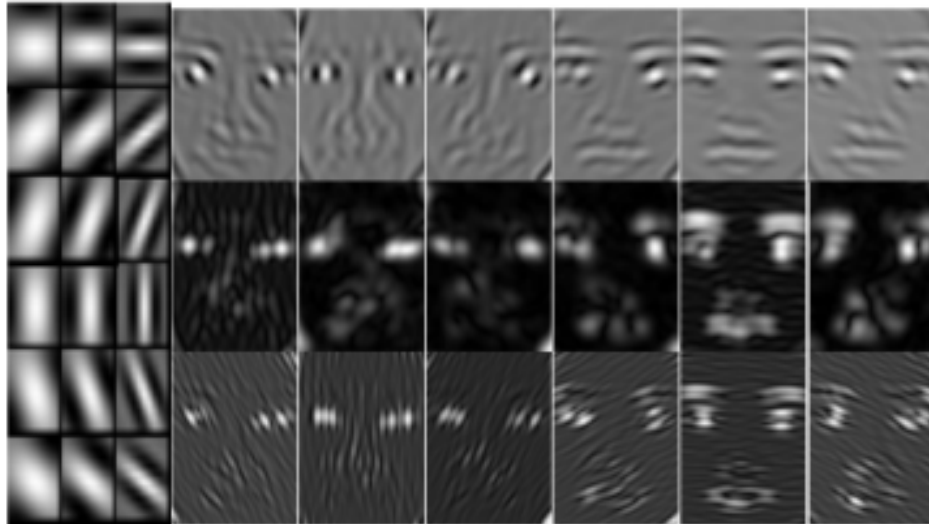


Рис. 3.38 Фильтры Габора и результат их применения локальные признаки изображения.

3.7.3.3. Методы распознавания на основе Скрытых марковских моделей

Скрытые марковские модели (СММ — статистическая модель, имитирующая работу процесса, похожего на марковский с неизвестными параметрами; ее задачей является разгадывание неизвестных параметров на основе наблюдаемых) [57] – это статистическая модель, используемая для описания статистических характеристик сигнала. Каждая модель $\lambda = (A, B, \pi)$ представляет собой набор N состояний $S = \{S_1, S_2, \dots, S_N\}$ между которыми возможны переходы.

Встроенные Скрытые марковские модели. В работе используются встроенные СММ, в которых каждое состояние представлено суперсостоянием, имеющим внутри себя набор состояний. В модели пять суперсостояний, соединенных в линейную СММ, количество встроенных состояний 3-6-6-6-3, соответствующих последовательности лоб – глаза – нос – рот – подбородок, внутренние состояния в каждом суперсостоянии объединены также в линейную СММ:

Внешняя модель задается следующими параметрами:

$$S_0 = \{S_{0i}, i = 1, 2, \dots, N_0\}$$

$$\pi_0 = \{ \pi_{0i}, i = 1, 2, \dots, N_0\}$$

$$A_0 = \{a_{0ij}, i, j = 1, 2, \dots, N_0\} \quad (4.26)$$

где S_0 – набор суперсостояний, π_0 – определяет начальное состояние, A_0 – вероятность переходов по суперсостояниям.

Каждая внутренняя модель задается следующими соотношениями:

$$\begin{aligned} S_1^k &= \{S_{1i}^k, i = 1, 2, \dots, N_1^k\} \\ \Pi_1^k &= \{\pi_{1i}^k, i = 1, 2, \dots, N_1^k\} \\ A_1^k &= \{a_{1ij}^k, i, j = 1, 2, \dots, N_1^k\} \\ B^k &= \{b_j^k(O_{t0t1})\} \\ b_j^k(O_{t0t1}) &= \sum_{m=1}^M c_{im}^k N(O_{t0t1}, \mu_m^k, U_m^k) \\ \Lambda &= \{\Pi_1^k, A_1^k, B^k\} \end{aligned} \quad (4.27)$$

где N – гауссиан, μ – вектор средних значений, U – матрица ковариации, $b_j^k(O_{t0t1})$ – мера близости наблюдения к эталонному, c – коэффициенты разложения по гауссианам (в работе использовалось разложение по 3-м гауссианам)

$$\Lambda = \{\Lambda^1, \Lambda^2, \dots, \Lambda^{N_0}\}$$

Модель СММ:

$$\Lambda = (\Pi_0, A_0, \Lambda)$$

1. Подготовка входных данных.

Из изображения выделяется последовательность наблюдений: изображение сканируется окном размерностью 16x16 слева направо сверху вниз с перекрытием 14 пикселей. Получается набор блоков.

Для каждого блока получаем коэффициенты дискретного косинусного разложения [13,40]:

$$C(u, v) = \frac{1}{4} w(u)w(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) \cos\left[\frac{\pi(2x+1)u}{MN}\right] \cos\left[\frac{\pi(2y+1)v}{MN}\right] \quad (3.28)$$

где x, y – пространственные координаты пикселя, $f(x, y)$ – значения пикселей исходного макроблока, u, v – координаты пикселя в частотном представлении, $w(u) = 1/\sqrt{2}$ при $u=0$, в остальных случаях $w(u) = 1$, $w(v) = 1/\sqrt{2}$ при $v = 0$, в остальных случаях $w(v) = 1$, $N=16$, $M = 16$ – ширина и высота окошка сканирования. Первые девять коэффициентов составляют входной вектор.

2. Построение Встроенной Скрытой марковской модели [57].

Алгоритм построения Встроенной Скрытой марковской модели:

- Равномерное разбиение входных векторов по состояниям;
- Инициализация модели;
- Применение разбиения Витерби для встроенной СММ. На этом этапе происходит перераспределение входных векторов по состояниям.
- Проводим оценку полученной модели, и либо возвращаемся к предыдущему шагу, либо говорим, что искомая модель построена.

Фактически происходит сегментация изображения: входные вектора разделяются на группы, т.е. каждый вектор относится к некоторому внутреннему состоянию. (см. рис. 3.39).

Недостатком СММ является то, что она не обладает различающей способностью. Т.е. алгоритм обучения только максимизирует отклик каждой модели на свои классы, но не минимизирует отклик на другие классы, и не выделяются ключевые признаки, отличающие один класс от другого. Для определения того, содержится ли лицо в обучающей выборке, использовался алгоритм ранжирования вероятностей, заключающийся в следующем. На обучающем наборе, каждая модель реагирует на изображения-примеры с некоторой вероятностью. Отсортированные таким образом модели образуют исходное ранжирование. Для неизвестного изображения модели так же ранжируются по вероятностям отклика на неизвестное изображение. Большая величина отклонения полученного ранжирования от исходного сигнализирует о том, что изображение принадлежит к неизвестному классу лица. Таким образом, похожие классы могут оказаться слабо различимыми и при увеличении объёма базы или использования в более широких условиях СММ может оказаться ненадёжными

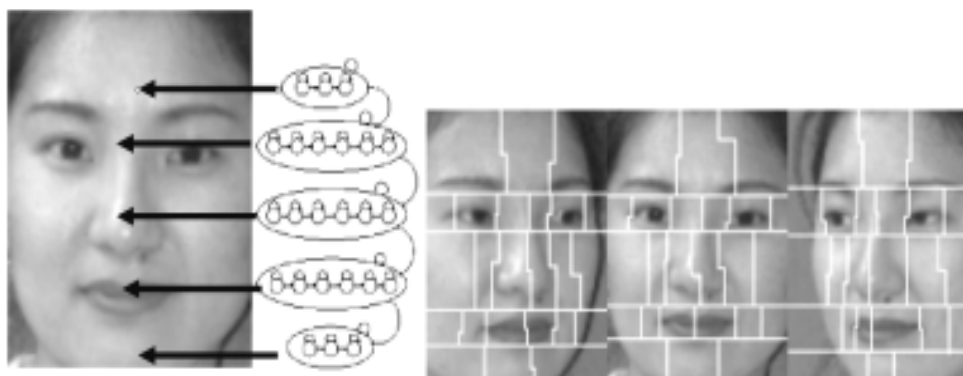


Рис. 3.39 Схема сегментации изображения на основе Скрытой марковской модели и Структурные признаки изображения, выделенные Скрытой марковской моделью.

Процесс распознавания в Скрытой марковской модели:

- Системе подается на вход некоторое изображение;
- Система строит из него входную последовательность;
- Система строит вероятности построения такой входной последовательности всеми моделями
- Выдается наиболее вероятный ответ и вероятность соответствия.

3.7.3.4. Результаты экспериментов по распознаванию лица

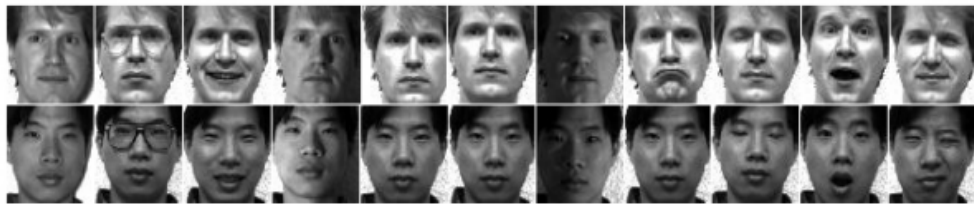
Рассмотрим результаты схожих экспериментов, проведенных в

различных исследовательских центрах на одинаковых наборах входных данных. Очень часто в исследованиях создаются свои базы изображений лиц из сети интернет, хотя в открытом доступе находятся различные базы изображений лиц, учитывающих различные особенности конкретной ситуации, такие как: освещенность, мимику, повороты головы, зашумленность, цветность, наличием \ отсутствием различных внешних объектов, как показано на рис. 3.40

В тестах использовались следующие выборки(см. рис. 3.41): обучающая выборка для 50 регистрируемых человек по 3 изображения в разных ракурсах для каждого (150 изображений), тестирующая выборка из 350 изображений (по 7 на каждого человека) и выборка из 1000 изображений незарегистрированных ни в одной базе лиц[49,90].



Образец лица из базы ORL



Образец лица из базы Yale



Образец лица из базы AR



Образец лица из базы MIT

Рис 3.40.Примеры изображений лиц из различных баз данных.

В ходе экспериментов применялись алгоритмы, как с использованием, так и без использования нейросетевых алгоритмов классификации. Результаты тестирования представлены в таблице 3.6

Предварительные результаты экспериментов над алгоритмами распознавания свидетельствуют[49]:

1) Алгоритм на основе метода главных компонент и нейронных сетей позволяет распознавать лицо даже при наличии значительных шумов и внешних объектов (очки). Тем не менее, по точности он уступает методу на основе локальных признаков и Марковских моделей. Это объясняется тем, что главные компоненты несут в себе главную информацию о признаках некоторого обобщенного лица, нежели чем позволяют определять уникальные признаки лица конкретного человека. [74,101].



Рис.3.41. Обучающая выборка и тестирующая выборка

2) Алгоритм на основе вейвлетов и нейронных сетей имеет наилучшую точность и позволяет наиболее точно определять уникальные локальные черты лица человека. Тем не менее, алгоритм более чувствителен к шумам на изображении, нежели метод главных компонент. [104,105]

3) Алгоритм на основе скрытых марковских моделей показывает точность сравнимую с анализом локальных признаков. Алгоритм позволяет определять структурные особенности лица и учитывать характер искажений лица. Данный метод более устойчив к шумам по сравнению с методом локальных признаков [52,57].

Результаты распознавания лиц

Метод	FAR (ошибочный допуск)	FRR (отказ в допуске)
Метод собственных фильтров [56]	8,6%	48,1%
Метод собственных фильтров + двухслойный перцептрон [56]	2,9%	36,5%
Метод собственных фильтров + сеть обратного распространения (Back Propagation Neural Network) [108]	3,5%	38,2%
Метод вейвлетов Габора + сеть обратного распространения (Back Propagation Neural Network) [120]	7,9%	37,3%
Метод вейвлетов Габора +двухслойный перцептрон[56]	2,9%	30%
Скрытая марковская модель[52]	2,9%	31%

3.8. Аппаратные и программные средства реализации ИНС

3.8.1. Программное обеспечение для моделирования ИНС

В настоящее время решение основной части практических задач с помощью ИНС обеспечивается использованием коммерческого программного обеспечения на основе методов нейробиологии, исполнение которых, что естественно, происходит с помощью центрального процессора. Это вполне логично, главным образом из-за того, что это относительно новая область, а создание ПО — достаточно гибкий процесс, что позволяет тестировать и внедрять с малыми затратами некоторые экспериментальные методы. Различают универсальные и прикладные программные продукты для моделирования нейронных сетей (Neural Network Simulators). Универсальные или объектно-инвариантные среды позволяют синтезировать оптимальные нейронные сети, применяемые для решения широкого класса задач, с предложением различных парадигм и алгоритмов обучения. Прикладные среды моделирования ориентированы для синтеза нейронных сетей, применяемых в той или иной отрасли, прикладной области или специфичной задаче. Среди важнейших свойств нейросетевых симуляторов – способность синтезировать код программы результирующей нейронной сети на алгоритмическом языке высокого уровня (чаще всего – С, С++, С#, Delphi). Такой код впоследствии легко интегрировать в пользовательскую программу. В таблице 3.7. представлены особенности наиболее распространенных универсальных программных сред для моделирования нейронных сетей [12,42,48].

Таблица 3.7.

Программные средства моделирования нейронных сетей

Наименование	Разработчик/производитель	Платформа	Поддерживаемые парадигмы и алгоритмы обучения	Интерфейс	Комментарии
1	2	3	4	5	6
MathLab Neural Network Toolbox	MathWorks, США, http://www.mathworks.com	Win 98, NT 4.0	Перцептрон, сети обратного распространения, сети Эльмана, ХопфильдаБ вероятностная и обобщенная регрессия	GUI для Ms Windows	Генерирует ANSI совместимый код.
SNNS	Институт параллельных и распределенных вычислений (IPVR) при Штутгартском университете, Германия, http://www.informatic.uni-stuttgart.de/ipvr/project/snns/annonce.html	Unix	Сети обратного распространения, сети радиального базиса, ART 1, ART 2, карты Кохонена, сети Эльмана, сети Джордана, ассоциативная память	GUI для X-Windows	Один из лучших симуляторов. Может работать с MS Windows при использовании симулятора X-Windows. Поставляются исходные коды на C++.
Trajan	Trajan Software Ltd., Великобритания, http://www.trajan-software.demon.co.uk/commerce.com	MS Windows	Сети обратного распространения, многоуровневый перцептрон, радиальный базис, карты Кохонена, вероятностная и обобщенная регрессия	GUI для Ms Windows	Стоимость - 620€, демо-версия доступна на сайте компании
Delta	Artificial Intelligence Group, Франция, Департамент компьютерных наук, http://www.inf.enst.fr/~milc/dnns.us.html	HP-UX Sun OS 4.1	Сети обратного распространения, многоуровневый перцептрон, карты Кохонена, сети Ельцина, сети Джордана	GUI для X-Windows	Демо-версия бесплатно
X-Sim	ИС, Испания, http://www.iic.uam.es/xsim/welcome.html	Unix, Linux	Сети обратного распространения, карты Кохонена, многоуровневый перцептрон, сети Джордана	Режим командной строки для OS Unix, GUI для X-Windows	Демо-версия бесплатно
Brain Wave	Университет Куинсланд, Австралия, http://www2.psy.uq.edu.au/~brainwav/	Любая	Сети обратного распространения, карты Кохонена, многоуровневый перцептрон, сети Хеббинена	Internet-браузер с поддержкой Java	Реализован в виде Java-апплет и может работать с любой ОС

1	2	3	4	5	6
VieNet2	Австрийский Институт исследования проблем искусственного интеллекта, http://www.wai.inivie.ac.at/oefai/nn/tool.html	DOS, WIN, Unix, Linux	Сети обратного распространения, сети радиального базиса, карты Кохонена, сети Эльмана, сети Джордана, ассоциативная память	Формируется пользователем	Распространяется в форме исходных кодов, что позволяет удобно использовать его для написания собственных программ
Neuro Windows	НейроПроект, Россия, http://neuroproject.ru	Win	Сети обратного распространения, карты Кохонена, ассоциативная память	Формируется пользователем	Библиотеки динамической компоновки Visual Basic, C++, Delphi - Стоимость - 450\$
Aspirin MIGRAINERS	Mitre Corp.	Unix	Сети обратного распространения	GUI для X-Windows	Обеспечивает сохранение весов и вектора узлов нейросети на диске в доступном формате
Atree	Билл Армстронг, Университет г. Альберта, Канада alnl@ualberta.ca	DOS, Unix	Адаптивные логические деревья	Режим командной строки для Unix-подобных ОС, окна для DOS	Демо-версия бесплатно
Snaps	Adaptive Solutions Inc.	Sun OS	Сети обратного распространения, карты Кохонена, LVQ2, частотно-чувствительное обучение	GUI для X-Windows	6875\$
ICSIM	Международный институт компьютерных наук, Беркли, США	Unix	Предопределенные сети	Shell, GUI для X-Windows	Демо-версия бесплатно
Neural Shell	Лаборатория SPANN, Департамент инженерной энергетики, Университет Огайо, США ftp://ftp.quanta.eng.ohio-state.edu/	Unix	Сети обратного распространения, сети Хопфильда, сети Хемминга, карты Кохонена, адаптивное медленное обратное распространение, частотно-чувствительное обучение	Режим командной строки для OS Unix, GUI для X-Windows	Демо-версия бесплатно
Neuron	Университет Дьюка, США	Unix	Трехмерная реконструируемая пирамидальная сеть	GUI для X-Windows	Демо-версия бесплатно
Sankom	Дортмундский университет, Германия	Unix	карты Кохонена	Режим командной строки Shell	Демо-версия бесплатно 188

1	2	3	4	5	6
SOMPA K	SOM, Лаборатория компьютерных и информационны х наук, Хельсингский университет технологий, Финляндия	DOS, Unix	Самоорганизующиеся карты и сети	Формируется пользователем	Демо-версия бесплатно
Xerion	Университет Торонто, Департамент компьютерных наук, Канада, ftp://ftp.cs.toront o.edu/pub/xerion	Unix	Сети обратного распространения, сети рекуррентного обратного распространения. Машина Больцмана, теория среднего поля, жесткое и мягкое конкурентное обучение, карты Кохонена	GUI для X- Windows	Демо-версия бесплатно
NETS	COSMIC, Университет Джорджи, США	DOS, Unix	Сети обратного распространения	Режим командной строки	Демо-версия бесплатно

Поскольку до сих пор большинство программных продуктов существуют только в ограниченных по функциональности демо-версиях, наиболее рациональным является рассмотрение пусть и не дешевого, но многофункционального симулятора Trajan.

Программный продукт TRAJAN компании TRAJAN Software Co. (Великобритания) является симулятором полного цикла и предназначен для моделирования в среде Windows различных видов нейронных сетей и алгоритмов обучения. Симулятор включает широкие возможности для графического и статистического контроля в процессе моделирования параметров и характеристик эффективности синтезируемых нейронных сетей. Рассмотрим основные функциональные и потребительские характеристики программного продукта TRAJAN. Программное обеспечение TRAJAN позволяет автоматизировать следующие процессы:

- формирования представительских выборок и переменных;
- выбора вида нейронной сети и ее структуры;
- обучения нейронной сети;
- сохранения оптимальных параметров нейронных сетей;
- выдачи графической и статистической информации о характеристиках синтезируемой нейронной сети.

Эти возможности позволяют работать в программной среде пользователям незнакомым с нюансами теории нейронных сетей, в то время как специалисты в данной области могут существенно сокращать рутинные этапы синтеза оптимальных нейронных сетей и временные затраты. При проектировании нейронной сети,

направленной на решение особенно сложных задач, моделирование в среде программного продукта TRAJAN, как правило, занимает несколько дней. Удобный пользовательский интерфейс программного продукта позволяет обеспечивать простой доступ к большим объемам информации. Представительские выборки и структура нейронной сети сохраняются в файлах с единым именем и различными расширениями, что обеспечивает легкость группировки исходных данных для моделирования. Обновляющиеся в реальном времени графики и гистограммы позволяют наблюдать за обучением и исполнением нейронной сети, оперативно реагировать на ход моделирования. При решении задач классификации или аппроксимации автоматически вычисляются разнообразные статистические параметры и характеристики. Специализированные топологические карты и кластерные диаграммы применяются при изучении и анализе результатов моделирования. Фактически вся символьная и числовая информация доступна в электронных таблицах (Datasheets), т.е. может быть мгновенно импортирована и экспортирована через буфер обмена Windows. Графическая информация также может быть экспортирована, например, для составления отчета о модельных экспериментах. Однако существуют и существенные ограничения, такие как:

- TRAJAN поддерживает нейронные сети «глубиной» 128 слоев, хотя, в подавляющем большинстве случаев, требуемое количество слоев существенно меньше.
- в TRAJAN первый слой является всегда слоем входа. Он используется только, чтобы вводить величины в нейронную сеть, так как нейроны входного слоя не подразумевают никакой обработки. Последний слой является выходным, и результаты выполнения нейронов этого слоя являются и выходом нейронной сети в целом.

3.8.2. Аппаратная реализация нейронных сетей

Не смотря на простоту и дешевизну программной реализации ИНС специализированные аппаратные средства (которые могут обеспечить поддержку или замену определенных программных пакетов) предлагают заметные преимущества. Рассмотрим глубже достоинства аппаратной реализации ИНС перед программным исполнением. К наиболее важным причинам следует отнести [9,48]:

- **Скорость.** Скорость исполнения большинства приложений может быть увеличена благодаря возможности проведения параллельных вычислений (особенно в задачах, требующих проведения повторных, итеративных вычислений). Благодаря чему снижается нагрузка на центральный процессор (если устройство

работает в составе системы), а, соответственно, повышается скорость исполнения других приложений.

- **Стоимость.** Применение устройств на основе ИНС может снизить общую стоимость системы. Как мы уже отметили выше, рассматриваемые аппаратные средства менее чувствительны к мощности ЦП, сохраняя при этом высокие показатели скорости. Этот фактор особенно важен в случае обновления или инсталляции новой крупной системы.

- **Надежность.** По схожим причинам аппаратная реализация может обеспечить более высокую надежность функционирования системы, в смысле меньшей вероятности отказа оборудования.

- **Специальные эксплуатационные режимы.** В ряде применений, налагающих ограничения на размеры, вес и другие физические характеристики, этот фактор может стать решающим.

- **Безопасность.** В плане защиты авторских прав, сопротивления взлому и другим противоправным действиям, применение аппаратной реализации позволяет обеспечить лучшую защиту по сравнению с эквивалентными функциями ПО на основе ИНС [9, 48].

Обобщенная классификация нейрочипов приведена на рис. 3.42.

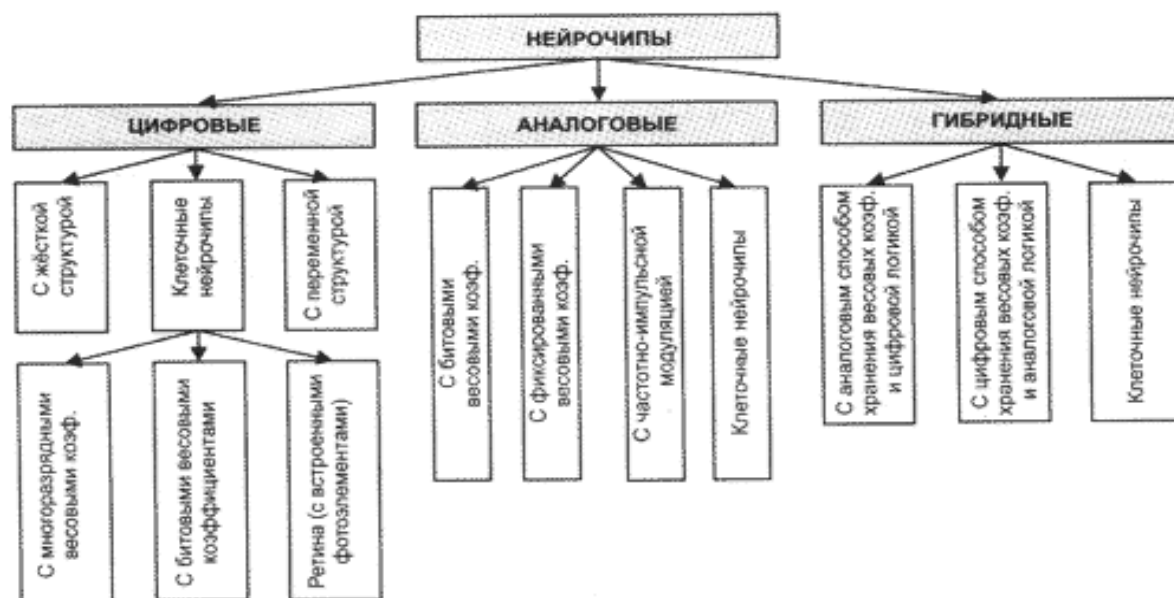


Рис. 3.42. Обобщенная классификация нейрочипов.

Основные характеристики и особенности аппаратных средств реализации ИНС приведены в таблице 3.8.[48].

Таблица 3.8.

Аппаратные средства реализации ИНС

Наименование	Фирма изготовитель	Разрядность, бит	Максимальное количество синапсов*	Максимальное число слоев**	Примечание
MA16	Siemens	48 (умножители и сумматоры)	-	-	400 ММАС.
NNP (Neural Networks Processor)	Accurate Automation	Nx16	-	-	MIMD, N - число процессоров.
CNAPS-1064	Adaptive Solutions	16	128 Кбайт	64	
100 NAP Chip	HNC	32	512 Кбайт	4	Плав. Арифм. 4 процессорных элемента
Neuro Matrix NM6403, Такт. частота 50 МГц.	Модуль, Россия	64 (вект. процессор), 32 RISC ядро	4096 шт.	24	Совместим с портами TMS320C4x
Neuro Matrix NM6404, Такт. частота 133 МГц.	Модуль, Россия	64 (вект. процессор), 32 RISC ядро	4096 шт.	~48	Совместим с портами TMS320C4x
CLNN 32 CLNN 64	Bellcore	32 64	496 1024	32 нейрона	10 ⁸ перекл./с 2 x 10 ⁸ перекл./с
NC 3001	NeuriGam	16	4096 шт.	32	
ZISC 036 (Zero Instruction Set Computer)	IBM	64 разр. входного вектора	-	36 нейронов	Частота 20МГц, Векторно-прототипный нейрочип
ETANN 80170NW	Intel	64 входа	Два банка весов 64x80	64 нейрона в слое, 3 слоя.	Аналоговая
MD-1220	Micro Devices	16	64 шт.	8	8 нейронов
MT 19003 - Neural Instruction Set Processor	Micro Circuit Engineering (MCE)	16 разр. Умножитель 35 разр. сумматор	-	1	RISC МП с 7 специальными командами
Neuro Fuzzu	National Semiconductor	-	-	-	
NI 1000	Nestor	5-16 (одного нейрона)	-	1024 прототипных 256 мерных векторов	Векторно-прототипный нейрочип
NLX420 (NLX 110, 230)	Adaptive Logic	16	1 Мбайт	16	16 процессорных элементов
OBL Chip	Oxford	16	16 Мбайт	-	

	Computer				
L-Neuro 1.0 L-Neuro 2.3	Philips	16 16	1536	16 нейронов 192 (12x16)	26 МГц 60 МГц
RSC (Speech Recognition Chip) - 164	Sensory Circuits	-	-	-	
ORC 110xx (Object Recognizer Chip)	Synaptics	-	-	-	
Pram-256 Chip	UCLi Ltd.	8 (одного нейрона)	-	256 нейронов	33МГц.
SAND	Datafactory	16	-	4	200 МСРС
ACC		16	-	-	
Геркулес	Россия	16	1 Мбайт	64	
Neuro Classifier	Университет Твента, DESY	70 вх. нейронов	-	6 (внутр) 1 вх., 1 вых.	2 x 1010 перекл./с
ANNA	AT&T	Число нейроннов 16-256	4096 весов	-	Число входов у нейрона 256-16.
WSC (Wafer Scale Integration)	Hitachi	-	64 связи на нейрон	576 нейронов	
SASLM2	Mitsubishi	2 (одного нейрона)	-	4096(64x64) нейронов	50 МГц
TOTEM	Kent (University UK), Trento (Italy)	16 (одного нейрона)	-	64 нейрона	30 МГц
Neuron 3120, Neurom 3150	Echelon (США)	8 бит (шина данных)	-	-	Наличие параллельных, последовательных и коммуникационных портов

* - максимальное число синапсов определяет размер внутрикристалльной памяти весов.

** - максимальное число слоёв определяется числом операций умножения с накоплением, выполняемых за один такт для операндов длиной 8 бит.

Для оценки производительности нейровычислителей используются следующие показатели:

- *CUPS* (connections update per second) - число измененных значений весов в секунду (оценивает скорость обучения).
- *CPS* (connections per second) - число соединений (умножений с накоплением) в секунду (оценивает производительность).
- $CPSPW = CPS/N_w$, где N_w - число синапсов в нейроне.

- $CPPS$ - число соединений примитивов в секунду, $CPPS = CPS * B_w * B_s$, где B_w , B_s - разрядность весов и синапсов.
- ММАС - миллионов умножений с накоплением в секунду.

Оценка производительности ряда процессоров при выполнении некоторых популярных алгоритмов ЦОС приведена в табл. 3.9.

Рассмотрим особенности современных нейропроцессоров на примере нейрочипа, разработанного в московском НТЦ "Модуль"[9]. За период с 1989 по 1999 гг. в НТЦ "Модуль" разработан ряд образцов нейрокомпьютерных вычислительных систем на современной элементной базе ведущих зарубежных фирм и на основе нейропроцессоров NeuroMatrix 6403 и 6404, созданного в НТЦ "Модуль". Среди широкой линейки различных ускорителей можно выделить: нейроускорители на основе универсальных сигнальных процессоров (в основном, используется TMS320C4x) и нейроускорители на базе специализированного нейросигнального процессора NeuroMatrix, созданного в НТЦ "Модуль"[9].

Таблица 3.9.

Производительность ряда процессоров при выполнении алгоритмов ЦОС

Наименование теста	Intel Pentium II, 300 МГц	Intel Pentium MMX, 200 МГц	Texas Instruments TMS320C40, 50 МГц	НТЦ "Модуль" NM6403, 40 МГц
Фильтр Собеля (размер кадра 384 × 288 байт), кадров/с.	-	21	6,8	68
Быстрое преобразование Фурье (256 точек, 32 разр.), мкс (тактов)	200	-	464 (11588)	102 (4070)
Преобразование Уолша-Адамара (21 шаг, входные данные – 5 бит), с	2,58	2,80	-	0,45

Сравнительные характеристики некоторых нейрочипов представлены в таблице 3.10.

Основными принципами, которыми руководствовались разработчики при создании архитектуры нейрочипа, были следующие:

- — программируемость операционных узлов, специально разработанных для эффективной поддержки нейровычислений;
- — необходимость обеспечивать вычисление значений выходов нейронов как можно большего участка нейросети за один процессорный такт;
- — необходимость производить вычисления над данным (синапсами и весами) произвольной разрядности (от 1 до 64 бит);
- — наличие в составе нейрочипа средств комплексирования для

создания распределенных параллельных вычислительных систем, состоящих из любого количества процессорных узлов;

- — наличие достаточно развитой системы команд, способной поддерживать наиболее часто используемые конструкции современных языков высокого уровня: любая команда (кроме команд загрузки выгрузки массивов данных) должна выполняться за один процессорный такт.

Таблица 3.10.

Сравнительные характеристики нейрочипов

Наименование нейрочипа	Конфигурация	CPS	CPSPW	CPPS	CUPS
NLX420	32-16, 8 bit mode	10M	20K	640M	-
100 NAP	4 chips, 2 M wts, 16 bit mantissa	250M	125	256G	64M
WSI (Hitachi)	576 neuron Hopfield	138M	3.7	10G	-
N64000 (Inova)	64-64-1, 8 bit mode	871M	128K	56G	220M
MA16	1 chip, 25 MHz	400M	15M	103G	-
ZISC036	64 8 bit element inp. Vector	-	-	-	-
MT19003	4-4-1-, 32 MHz	32M	32M	6.8G	-
MD1220	8-8	9M	1M	142M	-
NI 1000	256 5 bit element inp. Vector	40 000 vec in sec.	-	-	-
L-neuro-1	1-chip, 8 bit mode	26M	26K	1.6G	32M
NM6403	8 bit mode, 50 MHz	1200M	150M	77G	-

* В таблице приведены средние округленные показатели производительности

Рассмотрим принятые разработчиками архитектурные решения, удовлетворяющие приведенным выше требованиям. Взвешивание и сложение синапсов является наиболее сложной операцией и требует значительных аппаратных и временных затрат. Поэтому на данную операцию как основную ориентирован операционный узел нейрочипа, структурная организация которого позволяет выполнить взвешивание и сложение синапсов методом параллельного умножения их на весовые коэффициенты.

Операционный узел (ОУ) за один такт осуществляет взвешивание нескольких синапсов и вычисление их сумм одновременно для нескольких нейронов, т. е. за один такт ОУ реализует функции фрагмента нейросети, включающего в себя несколько синапсов и несколько нейронов.

Предварительно в ОУ загружаются соответствующие битовые элементы матрицы весов, содержимое которой определяется на этапе обучения нейросети. В данном режиме работы элементами матрицы являются разряды весовых коэффициентов и нулевые элементы, расположенные в строго фиксированных позициях. Такая настройка как бы разбивает ОУ на подматрицы размером $P \times R$, где P — количество нейронов, R — количество синапсов в реализуемом фрагменте нейросети. Пусть подматрица с координатами (p, r) соответствует p -му, $p=1, \dots, P$, нейрону и r -му, $r=1, \dots, R$, синапсу. Тогда каждая подматрица с координатами (p, l) вычисляет функцию $X_l W_{pl} + U_p$, а каждая подматрица с координатами (p, t) формирует произведение $X_t W_{pt}$, и прибавляет к нему результат, полученный на выходе подматрицы с координатами $(p, t-1)$, где $t=2, \dots, R$. Таким образом, на выходе подматрицы (p, R) сформируется результат операции.

$$Y_p = U_p + \sum_{r=1}^R X_r W_{pr} \quad (3.29.)$$

Необходимым условием работоспособности ОУ в данном режиме является следующее: количество бит, отводимых для каждого p -го нейрона, не должно превышать максимально возможную разрядность Y_p . Только в этом случае исключается возможность прохождения сигналов переноса, возникающих в цепях сложения, через границы между соседними нейронами. С одной стороны, данное условие обеспечивает объединение нескольких нейронов в единой вычислительной среде (каковой и является ОУ), а с другой — оно исключает возможность арифметических переполнений и, следовательно, необходимость дополнительного контроля. Выполнение этого условия должно обеспечиваться на этапе обучения нейросети путем введения в матрицу весов нулевых столбцов в области старших разрядов каждого нейрона.

Таким образом, количество нейронов, одновременно обслуживаемых ОУ, зависит от числа синапсов, их разрядности и значений весовых коэффициентов. Суммарная разрядность Y_p не должна превышать значения m . Количество синапсов, одновременно обрабатываемых ОУ, определяется разрядностью синапсов. Суммарная разрядность X не должна превышать значения m . Настройка нейрончика на работу с конкретными величинами разрядности синапсов и весов производится путем загрузки в специальные регистры управляющей информации как в начале процесса, так и в ходе его выполнения. Это позволяет динамически изменять разрядность обрабатываемых данных, а следовательно, скорость и точность вычислений от слоя к слою, от нейрона к нейрону, от синапса к синапсу.

При работе нейрочипа с одинаковой 8-битовой разрядностью синапсов и весов ОУ за один процессорный такт производит 24 операции умножения с накоплением, что при частоте тактовых импульсов 30 МГц составит 720 МСРС. Другими словами, за один процессорный такт производится вычисление участка нейросети, состоящего из 3 нейронов по 8 синапсов каждый, при этом разрядности нейронов и синапсов составляют 1 байт.

Общая структура нейрочипа приведена на рис. 3.42. (описание блоков и сигналов на схеме приведено на английском языке в силу устоявшихся традиций среди разработчиков микропроцессорных систем).

В состав нейрочипа входят:

- Vector processor (векторный процессор) — предназначен для реализации базовых операций в нейросетях над массивами данных, представленных в виде векторов с элементами, которые могут иметь переменную разрядность: основным узлом векторного процессора является операционный узел, описанный выше;
- Scalar processor (скалярный процессор) — предназначен для вычисления адресов команд, управления их выборкой, вычисления адресов операндов и весовых коэффициентов при работе нейрочипа с памятью, а также для поддержки скалярных операций над данными;
- MUX — два идентичных программируемых интерфейса с локальной и глобальной шинами — предназначены для работы с двумя внешними памятьями (SRAM и DRAM) по 64-разрядным шинам данных; тип памяти определяется в процессе инициализации;
- DMA (ПДП) — два коммуникационных порта с сопроцессорами ПДП, предназначенные для поддержки доступа в память при обмене по двунаправленным байтовым линкам (Link 0 и Link 1) через коммуникационные порты (Port 0 и Port 1) на фоне вычислений;
- Control unit (устройство управления) — выполняет функции общего управления и осуществляет арбитраж использования внешней памяти.

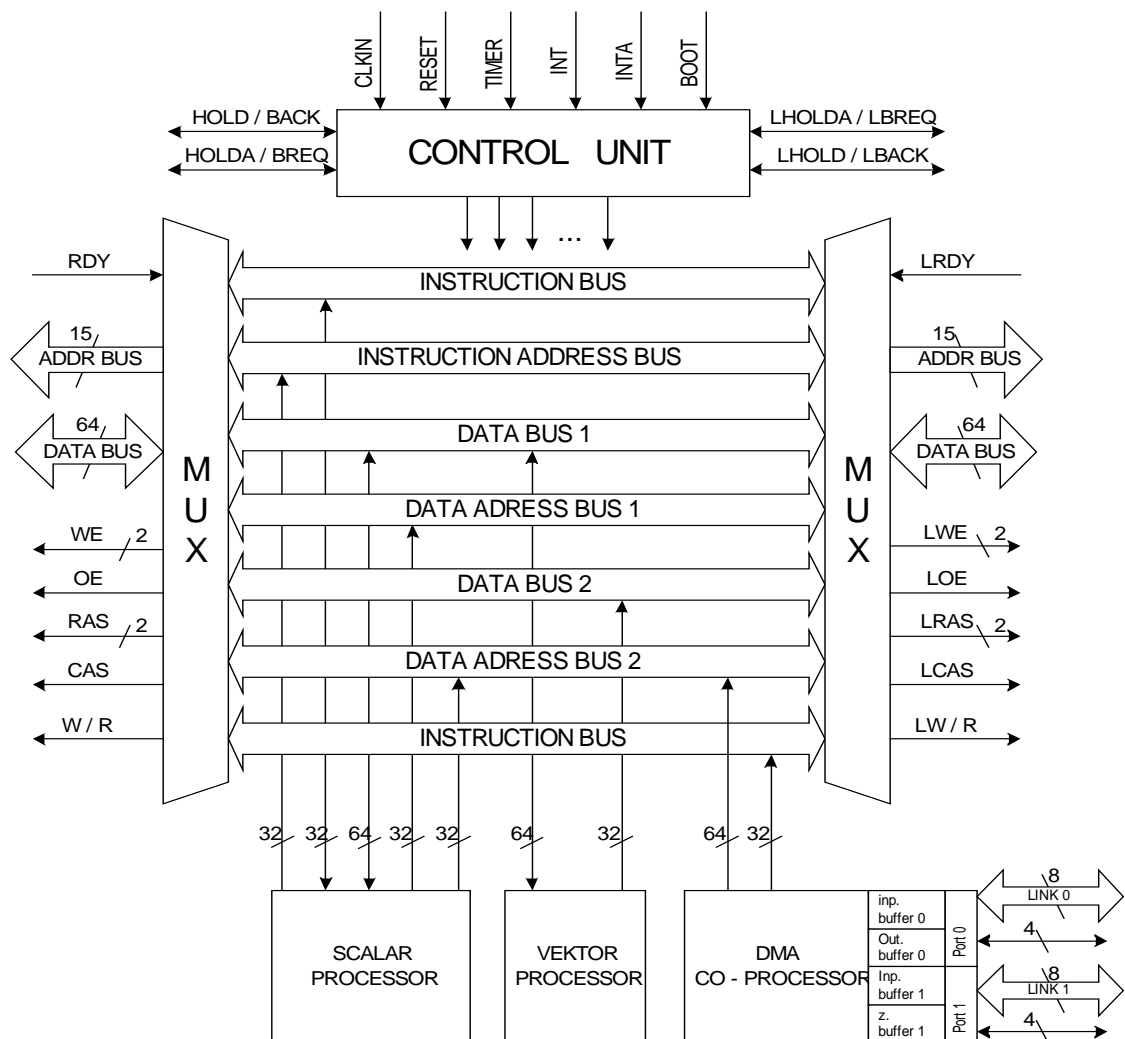


Рис. 3.42. Структура нейрочипа.

Нейрочип использует следующие внутренние шины:

- Instruction bus — 32-разрядную шину команд, по которой выбираются команды из памяти;
- Instruction address bus — 32-разрядную шину адреса команд, по которой выставляется адрес выбираемой команды;
- Data bus 1, 2 — 64-разрядные шины данных, обеспечивающие возможность одновременной работы с операндами и подкачку весовых коэффициентов в векторный процессор;
- Data address bus 1, 2 — 32-разрядные шины адреса данных;
- DMA bus — 64-разрядную шину ПДП, предназначенную для обмена в режиме ПДП с линками;
- DMA address bus — 32-разрядную шину адреса ПДП.

Шины памяти имеют сквозную адресацию. Старшим разрядом вычисленного адреса определяется, в какую из двух шин памяти производится доступ. Тем самым обеспечивается возможность получения данных и команд из любой памяти. Обмен с каждой из памяти в зависимости от команды может осуществляться либо словом, либо полусловом в соответствии с младшим разрядом

вычисленного адреса, который указывает, с какой из половин 64-разрядной памяти необходимо работать.

Нейрочип работает с командами фиксированной длины по 32 разряда, которые можно разделить на 4 основные группы: команды скалярного процессора по обработке операндов, команды векторного процессора, команды управления и команды пересылки.

Система команд скалярного процессора организована таким образом, что базовые операции (т. е. операции над регистрами общего назначения) могут совмещаться с операциями над адресными регистрами.

3.8.3. Принципы построения нейросистем на базе нейрочипа

Одной из особенностей нейросетевых методов обработки информации является высокая параллельность вычислений и, следовательно, целесообразность использования специальных средств аппаратной поддержки. В значительной мере успех в решении рассмотренных задач обусловлен использованием оригинальных ускорительных плат. Такие платы работают параллельно с процессором обыкновенного ПК и несут на себе основную вычислительную нагрузку, превращая основной процессор компьютера в устройство управления и обслуживания мощных вычислительных средств, расположенных на ускорительной плате[9,48].

Основными архитектурными возможностями нейрочипа для построения различных параллельных систем являются наличие двух высокоскоростных двунаправленных байтовых коммуникационных портов, аппаратно совместимых с портами сигнального процессора TMS320C4x (Семейство процессоров TMS320C4x объединяет процессоры TMS320C40 с шестью линиями для внешних связей и процессоры TMS320C44 с четырьмя линиями для внешних связей), и поддержка доступа к совместно используемой памяти. Различные варианты объединения нейрочипов позволяют добиться реализации большого числа высокопроизводительных параллельных систем разнообразной конфигурации. На рис.3.43 приведены примеры построения вычислительных сетей на базе нейрочипа: рис. 3.43,*а* — двунаправленный конвейер (для операций над матрицами и эмуляции нейросетей прямого распространения и других конвейеризированных вычислений); рис. 3.43,*б* — структура типа двухмерной решетки (для операций над матрицами и эмуляции нейросетей прямого распространения/

Кроме перечисленных методов можно также создавать вычислительные сети практически любой конфигурации с

использованием сигнального процессора TMS320C4x в качестве коммутирующего элемента.

Примеры такого подхода приведены на рис. 3.44 — структура типа дерева (для эмуляции многослойных нейросетей, а также для задач распознавания образа).

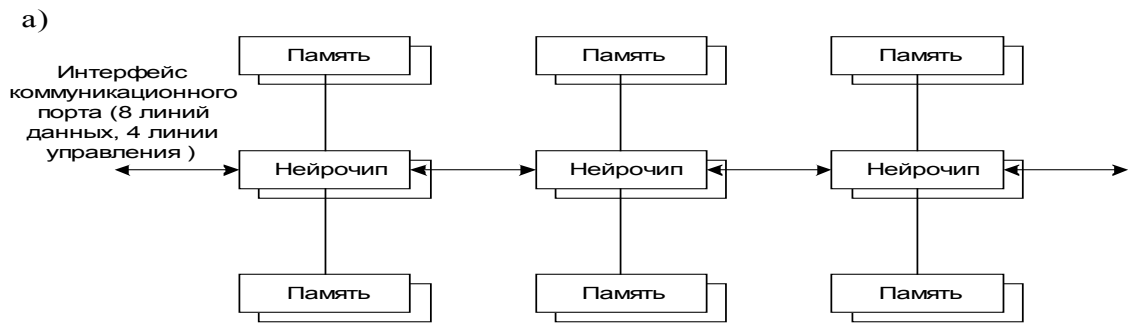
Используемый интерфейс с памятью определяет следующие группы архитектур систем на базе нейрочипа (рис. 3.45):

- — архитектура с совместно используемой памятью (глобальная память доступна нескольким нейрочипам) — рис. 3.45, *а*);

- — архитектура с распределенной памятью (каждый нейрочип имеет свою локальную память, а взаимодействие нейрочипов осуществляется через коммуникационные порты) — рис. 3.45, *б*);

- смешанная архитектура (каждый нейрочип имеет свою локальную память, а также возможность доступа к глобальной памяти вместе с другими нейрочипами) — рис. 3.45, *в*).

При объединении небольшого количества нейрочипов в систему можно применить архитектуру с единой глобальной памятью, но если число используемых нейрочипов большое, затраты на доступ в глобальную память становятся слишком велики, поэтому рекомендуется архитектура с распределенной памятью.



б)

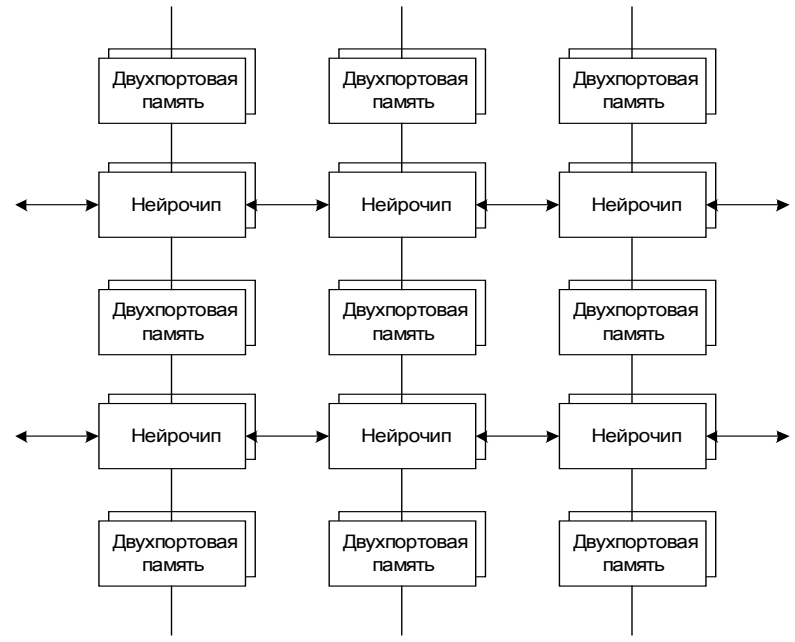


Рис. 4.43. Примеры архитектур нейросетей на базе нейрочипов.

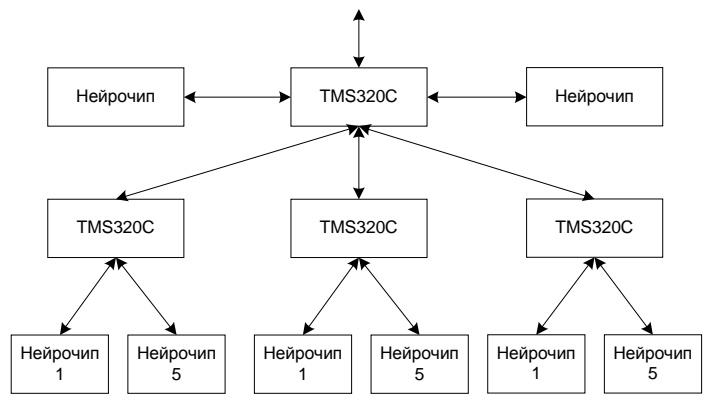
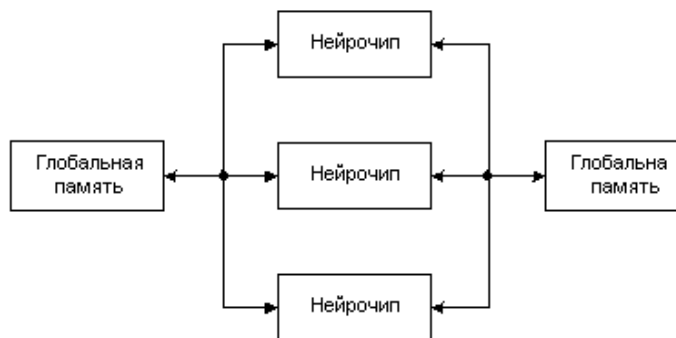
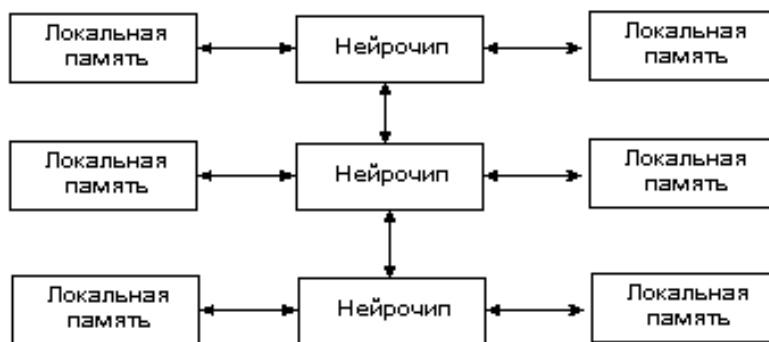


Рис.3.44. Древовидная структура ИНС

а)



б)



в)

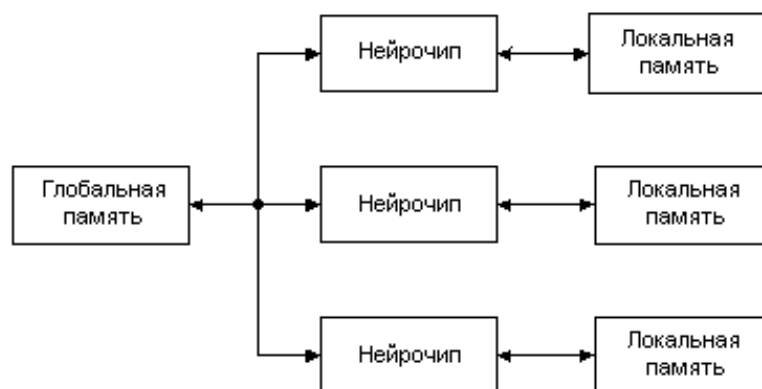


Рис. 3.45. Архитектуры ИНС с памятью

Заключение

Основным достоинством применения нейронных сетей является возможность решать различные неформализованные задачи. При этом можно очень просто моделировать различные ситуации, подавая на вход сети различные данные и оценивая выдаваемый сетью результат.

В ходе применения нейронных сетей отмечен существенный недостаток: сложность понимания процесса получения сетью результата. Первым шагом к устранению данной проблемы является разработка новой технологии, которая позволяет генерировать описание процесса решения задачи нейронной сетью. Используя таблицу экспериментальных данных, описывающих предметную область, можно будет получить явный алгоритм решения поставленной задачи. Требуется дальнейшие эксперименты по исследованию алгоритмов обучения, выбору начального представления и внедрению в архитектуру НС учёта свойств изображения.

Хотя существует множество аппаратных и программных средств моделирования нейросетей до сих пор чаще всего используются модели создаваемые на языках высокого уровня под каждый конкретный эксперимент что приводит к неоправданно высоким затратам и низкому быстродействию систем. Также часто используются свои собственные базы данных лиц и дикторов, что существенно усложняет процесс сравнения результатов различных экспериментов, хотя в открытом доступе есть специализированные биометрические базы (например, лиц с различными особенностями, речи). Практически полностью отсутствуют эксперименты с аппаратными реализациями нейросетевых методов биометрического распознавания.

Одномодальные системы биометрической идентификации не всегда обеспечивают необходимую надежность идентификации. Для обеспечения повышенной точности и устойчивости к НСД необходимо объединять несколько одномодальных систем в многомодальную на нескольких уровнях объединения (например слияние одномодальных систем идентификации по голосу, лицу и отпечатку пальца позволяет получить почти 100% точность идентификации) [18,38].

ЛИТЕРАТУРА

1. Абламейко С.В., Лагуновский Д.М. Обработка изображений: технология, методы, применение.//Беларусь, – Минск: Ин-т техн. Кибернетики НАН Беларуси, 2000, – 304 с.
2. Блаттер К. Вейвлет-анализ. Основы теории.// – М. – Техносфера. - 2006, - 279 с.
3. Вапник В.Н., Червоненкис А. Теория распознавания образов// М., Наука, 1974, - 415 с.
4. Васильев В.Н, Гуров И.П., Потапов А.С. Математические методы и алгоритмическое обеспечение анализа и распознавания изображений в информационно-телекоммуникационных системах//Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы», - 2008, - 46 с. <http://www.ict.edu.ru/itkonkurs2008/>
5. Васильев В.Н., Гуров И.П., Потапов А.С. Современная видеоинформатика: проблемы и перспективы// птический журнал, 2012, № 11, с. 5-15.
6. Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности//Украина, – Харьков: Основа, 1997. (<http://www.neuropower.de/rus>).
7. Востриков А.С., Пустовой Н.В. Цифровая обработка изображений в информационных системах.//Учебник НГТУ, Новосибирск 2002.
8. Галушкин А. И., Томашевич Д. С., Томашевич Н. С. Методы реализации инвариантности к аффинным преобразованиям двумерных изображений // Приложение к журналу «Информационные технологии», – 2001, – №1, с. 1-19.
9. Галушкин А.И. Некоторые исторические аспекты развития элементной базы вычислительных систем с массовым параллелизмом (80- и 90- годы) //Нейрокомпьютер, №1, 2000, с. 68-82.
10. Глазунов А. Компьютерное распознавание человеческих лиц//Открытые системы, 2000, №3, (<http://www.osp.ru/os/2000/03/>).
11. Головкин В.А. Нейроинтеллект: Теория и применения. Книга 1. Организация и обучение нейронных сетей с прямыми и обратными связями//Беларусь,. – Брест: БПИ, 1999, – 260с.
12. Головкин В.А. Нейроинтеллект: Теория и применения. Книга 2. Самоорганизация, отказоустойчивость и применение нейронных сетей. //Беларусь, – Брест: БПИ, 1999, – 228с.
13. Гонсалес Р., Вудс Р. Цифровая обработка изображений.// Пер. с англ.- Москва.- Техносфера. – 2006, -1072 с.
14. Горбань А., Россиев Д. Нейронные сети на персональном компьютере// Новосибирск: Наука, 1996, - 276 с.

15. Горелик А.Л., Скрипкин В.А. Методы распознавания//– М: Высшая школа, 1984, – 208 с.
16. Дуда Р., Харт П. Распознавание образов и анализ сцен// - М., Мир, 1976, - 511 с.
17. Кухарев Г.А. Биометрические системы: Методы и средства идентификации личности человека// - Спб., Политехника, 2003, - 240 с.
18. Кухарев Г. А. Методы обработки и распознавания изображений лиц в задачах биометрии / Г. А. Кухарев, Е. И. Каменская, Ю. Н. Матвеев, Н. Л. Щеголева; под ред. М. В. Хитрова. — СПб.: Политехника, 2013, — 388 с.
19. Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов// - М., Радио и связь, 1987, - 400 с.
20. Осовский С. Нейронные сети для обработки информации// — М.: Финансы и статистика, 2002
21. Панканти Ш, Боле Р.М., Джейн Э. Биометрия: будущее идентификации // Открытые системы, 2000, №3, (<http://www.osp.ru/os/2000/03/>).
22. Пентланд А.С., Чаудхари Т. Распознавание лиц для интеллектуальных сред // Открытые системы, 2000, №3, (<http://www.osp.ru/os/2000/03/>).
23. Потапов А.С. Распознавание образов и машинное восприятие// – СПб., Политехника, 2007, - 548 с.
24. Потапов А.С. Исследование представлений изображений на основе принципа репрезентационной длины описания // Изв. вузов. Приборостроение, 2008, Т. 51, № 7, с. 3-7.
25. Потапов А.С., Малышев И.А., Луцив В.Р. Совмещение аэрокосмических изображений с субпиксельной точностью методом локальной корреляции//Оптический журнал, 2004, Т. 71, № 5, с. 31–36.
26. Прэтт У. Цифровая обработка изображений// - М., Мир, 1982, Кн. 2, - 480 с.
27. Разин И.В., Тетерин В.В. Математическая модель для экспресс-оценки степени сходства изображений//Оптический журнал, 2001, № 11, с. 33–37.
28. Рубина И. С., Тропченко А. Ю. Исследование алгоритмов выбора опорных пикселей в задачах выделения сегментов кадра видеопоследовательности // Известия вузов. Приборостроение, - СПб., 2012, № 1, т.55, с. 9 – 14.
29. Садыхов Р.Х., Ваткин М.Е. Модифицированный алгоритм обучения РБФ-сети для распознавания рукописных символов // Идентификация образов. – Минск: Ин-т техн. Кибернетики НАН Беларуси, 2001, с.7-16.

30. Самаль Д.И. Построение систем идентификации личности на основе антропометрических точек лица // Цифровая обработка изображений. - Минск:ИТК, 1998, с.72-79.
31. Самаль Д.И., Старовойтов В.В. - Подходы и методы распознавания людей по фотопортретам. - Минск, ИТК НАНБ, 1998, - 54с.
32. Самаль Д.И., Старовойтов В.В. Выбор признаков для распознавания на основе статистических данных // Цифровая обработка изображений, – Минск, Ин-т техн. кибернетики НАН Беларуси, 1999, с.105-114.
33. Самаль Д.И., Старовойтов В.В. Методика автоматизированного распознавания людей по фотопортретам // Цифровая обработка изображений. – Минск:Ин-т техн. кибернетики НАН Беларуси, 1999, с.81-85.
34. Старовойтов В.В., Талёб М.А. – Методы сегментации цветных изображений// – Минск Ин-т техн. кибернетики НАН Беларуси, 1999, – 44с.
35. Самаль Д.И. Построение систем идентификации личности на основе антропометрических точек лица // Цифровая обработка изображений. - Минск:ИТК, 1998, с.72-79.
36. Талалаев А. А., Тищенко И. П., Фраленко В. П., Хачумов В. М. Анализ эффективности применения искусственных нейронных сетей для решения задач распознавания, сжатия и прогнозирования // Искусственный интеллект и принятие решений, № 2, 2008, с. 24–33
37. Тропченко А. А., Тропченко А. Ю. Нейросетевые методы идентификации человека по изображению лица // Известия вузов. Приборостроение, - СПб., 2012, № 10, Т.55, с. 31– 6.
38. Тропченко А. А. Методы повышения робастности распознавания в мультимодальных биометрических системах // Известия вузов. Приборостроение, - СПб., 2014. № 4, Т.57, с. 20–23.
39. Тропченко А.Ю., Тропченко А.А. «Цифровая обработка изображений методы сжатия изображений, аудио и видео данных» // - СПб., - СПбГУ ИТМО, 2009, -120с..
40. Тропченко А.Ю., Тропченко А.А. «Цифровая обработка сигналов. Методы предварительной обработки» // - СПб., - СПбГУ ИТМО, 2009, -90с.
41. Тропченко А. Ю., Ахунзянов Р. Р. Разработка адаптивного детектора тона кожи // Научно-технический вестник информационных технологий, механики и оптики, СПб., 2013, № 5, Т.87, с. 103–107.
42. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика//– М., МИФИ, 1992, – 184 с. (<http://www.neuropower.de/rus>).
43. Фурман Я.А. и др. Введение в контурный анализ и его приложения к обработке изображений и сигналов// -М., - ФИЗМАТЛИЗ, 2002, -592 с..

44. Фу К. Структурные методы в распознавании образов// - М., Мир, 1977, - 320 с.
45. Фор А. Восприятие и распознавание образов. // - М.; Машиностроение, -1989, - 272 с.
46. Хорн Б.К.П. Зрение роботов// – М., Мир, 1989, – 488 с.
47. Шапиро Л., Стокман Дж. Компьютерное зрение// – М., БИНОМ, - 2008.
48. Шахнов В.А., Власов А.И., Кузнецов А.С., Поляков Ю.А. Нейрокомпьютеры: архитектура и схемотехника.// - М., Изд-во Машиностроение, 2000. - 64 с.
49. Bazanov P, Tae-Kyun Kim, Seok Cheol Kee, Sang Uk Lee. Hybrid and Parallel Face Classifier based on artificial networks and principal component analysis// International Conference of Image Processing 2002 Rochester , New York September 22-25 2002, Vol. 1, pp 916-919
50. Blanz V. & Vetter T. (2003). Face Recognition Based on Fitting a 3D Morphable Model// IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, September 2003, pp. 1063–1074
51. Brunelli R., Poggio T. Face recognition: features versus templates // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, Vol. 15, № 10, pp. 235-241.
52. Bryliuk D., Starovoitov V. Application of Recirculation Neural Network and Principal Component Analysis for Face Recognition // The 2nd International Conference on Neural Networks and Artificial Intelligence, – Minsk, BSUIR, 2001, pp.136-142.
53. Belhumeur P. N., Hespanha J. P. and Kriegman D. J. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection// IEEE Transactions on Pattern Analysis and Machine Intelligence 1997, Vol. 19, pp. 711-720.
54. Chan T.F., Shen J., and Vese L. Variational PDE models in image processing // Notice Amer. Math. Soc., 2003, Vol 50, pp. 14–26.
55. Chellappa R, C.L. Wilson C. L., Sirohey S. Human and machine recognition of faces: A survey. //Proceedings of the IEEE, № 83, Vol.5, pp. 705–740, 1995.
56. Chen W.; Yuen P. C., Huang J. & Dai D., Kernel Machine-Based One-Parameter Regularized Fisher Discriminant Method for Face Recognition//IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 35, №. 4, August 2005, pp. 659-669.
57. Eickeler S., Muller S., Rigoll G. High performance face recognition using Pseudo 2-D Hidden Markov Models // Gerhard-Mercator-University Duisburg, Germany, 1998, - 6 p.
58. Eickeler S., Muller S., Rigoll G. Recognition of JPEG Compressed Face Images Based on Statistical Methods // Gerhard-Mercator-University Duisburg, Germany, 1999, - 17 p.

59. Er M. J.; Chen W. & Wu S. (2005). High-Speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks//IEEE Transactions on Neural Networks, Vol. 16, № 3, May 2005, pp. 679-698.
60. Esme B., Sankur B., Anarim E. Facial feature extraction using genetic algorithms // 8-th European Signal Processing Conference, Trieste, 1996, pp. 1511-1514.
61. Essannouni L., Ibn-Elhaj E., Aboutajdine D. Fast cross-spectral image registration using new robust correlation // Journal of Real-Time Image Processing, 2006, Vol. 1. № 2. pp. 123–129.
62. Galland F., Bertaux N., Réfrégier Ph. Multi-component image segmentation in homogeneous regions based on description length minimization: Application to speckle, Poisson and Bernoulli noise // Pattern Recognition, 2005, Vol. 38. Iss. 11. pp. 1926–1936.
63. Gao Y. & Leung M. K. H. (2002). Face Recognition Using Line Edge Map// IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 6, June 2002, pp. 764–779.
64. Goecke R., Asthana A., Pettersson N., Petersson L. Visual vehicle egomotion estimation using the Fourier-Mellin transform // IEEE Trans. Intelligent Vehicles Symposium, 2007, pp. 450–455.
65. Graham D. B., Allinson N. M. Face recognition using virtual parametric eigenspace signatures//Image Processing and its Applications, pp. 106-110, 1997.
66. Grudin M. A., Lisboa P. J., Harvey D. M. Compact multi-level representation of human faces for identification// Image Processing and its Applications, pp. 111-115, 1997.
67. Gutta S., Wechsler H. Face recognition using hybrid classifiers // Pattern Recognition 1997, Vol. 30, pp. 539-553.
68. Hafed Z. M. & Levine M. D. Face Recognition Using the Discrete Cosine Transform//International Journal of Computer Vision, Vol. 43, №. 3, July 2001, pp. 167–188.
69. Moghaddam B. and Pentland A. Probabilistic Visual Learning for Object Representation// IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, Vol. 19, pp. 696-710.
70. Hallinan P. L., Gordon G. G., Yuille A. L., Giblin P., Mumford D. Two- and Three-Dimensional Patterns of the Face// Natick:A, K. Peters Ltd., 1999.,- 260 p.
71. Hao Wu, Yunqiang Chen, Tong Fang. Coupled Bayesian framework for dual energy image registration // Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, Vol. 2, pp. 2475–2482.
72. Hörster E., Lienhart R., Slaney M. Image retrieval on large-scale image databases // Proc. 6th ACM int. conf. on Image and Video Retrieval, CIVR 2007, 2007, pp. 17–24.

73. Keyzers D., Deselaers T., Gollan C., Ney H. Deformation models for image recognition // *Trans. on Pattern Analysis and Machine Intelligence*, 2007, Vol. 29, № 8, pp. 1422–1435
74. Kong H.; Li X., Wang L., Teoh E. K., Wang J. & Venkateswarlu R. (2005). Generalized 2D Principal Component Analysis// *IEEE International Joint Conference on Neural Networks (IJCNN)*, Montreal, Canada, August 2005, pp.108-113.
75. Kopperapu S.K., Desai U.B. Bayesian approach to image interpretation // *The Springer International Series in Engineering and Computer Science*, 2001, Vol. 616. - 127 p.
76. Krueger N. An Algorithm for the Learning of Weights in Discrimination Functions Using a Priori Constraints// *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, Vol. 19, pp. 764-768
77. Kruizinga P., Petkov N. Optical flow applied to person identification // *Proceedings of the EUROSIM Conference on Massively Parallel Processing applications and Development*, 1994, pp. 871-878.
78. Kuchariew G., Forczmanski P. Hierarchical method of Reduction of Features Dimensionality for Image Recognition and Graphical Data Retrieval // *Pattern Recognition and Image Processing*, 2002, Vol. 1, pp. 57-72.
79. Kwak K. C. & Pedrycz W., Face Recognition Using Fuzzy Integral and Wavelet Decomposition Method, *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, Vol. 34, № 4, August 2004, pp. 1666–1675
80. Lillholm M., Nielsen M., Griffin L.D. Feature-Base Image Analysis // *Int. J. Computer Vision*. 2003, Vol. 52, №. 2/3, pp. 73–95
81. Lanitis A., Taylor C. J., Cootes T. F. Automatic Interpretation and Coding of Face Images Using Flexible Models// *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, Vol. 19, pp. 743-756
82. Lindeberg T. Edge detection and ridge detection with automatic scale selection // *Int. J. of Computer Vision*. 1998. V. 30. Iss. 2. P. 117–156.
83. Lowe D. Object recognition from local scale-invariant features // *Proc. Int. Conf. on Computer Vision*, 1999, pp. 1150–1157.
84. Luo Q., Khoshgoftaar T.M. Unsupervised multiscale color image segmentation based on MDL principle // *IEEE Trans. on Image Processing*. 2006, Vol. 15, № 9, pp. 2755–2761.
85. Nazeer, S. A.; Omar, N. & Khalid M., Face Recognition System using Artificial Neural Networks Approach// *International Conference on Signal Processing, Communications and Networking (ICSCN '07)*, 22-24 Feb. 2007, pp 420-425.
86. Nazeer, S. A.; Khalid M, PCA-ANN Face Recognition System based on Photometric Normalization Techniques// *State of the Art in Face Recognition*, January 2009, I-Tech, Vienna, Austria, pp. 71-86.

87. Popescu A., Moëllic P.-A., Millet Ch. SemRetriev: an ontology driven image retrieval system. // Proc. 6th ACM Int. conf. on Image and Video Retrieval. 2007, pp. 113–116.
88. Rabiner L. R. A tutorial on Hidden Markov Models and selected applications in speech recognition // Proceedings of the IEEE, 1989, Vol. 77, №2, -pp. 257-285.
89. Sadykhov R., Vatkin M. New training algorithm of neural network “Neokognitron” for a recognition of hand-written symbols // Pattern Recognition and Image Processing. – 2001. – P. 203-212.
90. Samaria F. Face Recognition Using Hidden Markov Models // PhD thesis, Engineering Department, Cambridge University, 1994.
91. Smeraldi, F., Bigun, J. Retinal Vision applied to facial features detection and face authentication // Pattern Recognition Letters, 2002, Vol. 23, №. 4, pp. 463–475
92. Sung K. K., Poggio T. Learning Human Face Detection in Cluttered Scene // Lecture Notes in Computer Science - Computer Analysis of Images and Patterns, 1995. pp. 432-439.
93. Tu Z.W. and Zhu S.C. Image Segmentation by data driven Markov chain Monte Carlo // IEEE Trans. PAMI, 2002, Vol. 24, №. 5, pp. 657–673.
94. Vetter T., Poggio T. Linear Object Classes and Image Synthesis From a Single Example Image // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, Vol. 19, pp. 733-742
95. Wang F., Vemuri B. Non-rigid multi-modal image registration using cross-cumulative residual entropy // Int. J. Computer Vision, 2007, Vol. 74, № 2, pp. 201–215..
96. Wiskott L., Fellous J.-M., Krueger N and Malsburg C. Face Recognition by Elastic Bunch Graph Matching // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, Vol. 19, pp. 775-779.
97. Wurtz R. P. Object Recognition Robust Under Translations, Deformations, and Changes in Background // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, Vol. 19, pp. 769-775.
98. Yan S.; He X., Hu Y., Zhang H., Li M., Cheng Q. Bayesian Shape Localization for Face Recognition Using Global and Local Textures // IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, № 1, January 2004, pp. 102-113
99. Yang C.H.T., Lai S.H., Chang L.W. Hybrid image matching combining Hausdorff distance with normalized gradient matching // Pattern Recognition, 2007, Vol. 40, № 4, pp. 1173–1181.
100. Zhao W., Chellappa R., Phillips P.J., Rosenfeld A. Face recognition: a literature survey // ACM Computing Surveys, 2003, Vol. 35, pp. 399–458.
101. Zavorin I., LeMoigne J. Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery // IEEE Trans. Image Processing, 2005, Vol. 14, № 6, pp. 770–782

102. Zheng W.; Lai J. & Yuen P. C., A New LDA-Based Face Recognition Algorithm With Selection of Principal Components, IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 35, No. 5, October 2005, pp. 1065– 1078
103. Zuo F. & de With P. H. N. Real-time Embedded Face Recognition for Smart Home//IEEE Transactions on Consumer Electronics, Vol. 51, № 1, February 2005, pp. 183-190
104. Zhang B.; Zhang H., Ge S. S., Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory// IEEE Transactions on Neural Networks, Vol. 15, № 1, January 2004, pp. 166–177.
105. Zhang H.; Zhang B., Huang W., Tian Q. Gabor Wavelet Associative Memory for Face Recognition// IEEE Transactions on Neural Networks, Vol. 16, № 1, January 2005, pp. 275–278.
106. Zhu S.C., Wu Y., Mumford D. Minimax entropy principle and its application to texture modeling // Neural Computation, 1997, № 9, pp. 1627–1660

Миссия университета – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для содействия решению актуальных задач.

КАФЕДРА ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ

О кафедре

Кафедра вычислительной техники Университета ИТМО создана в 1937 году и является одной из старейших и авторитетнейших научно-педагогических школ России.

Первоначально кафедра называлась кафедрой математических и счетно-решающих приборов и устройств и занималась разработкой электромеханических вычислительных устройств и приборов управления. Свое нынешнее название кафедра получила в 1963 году.

Кафедра вычислительной техники является одной из крупнейших в университете, на которой работают высококвалифицированные специалисты, в том числе 7 профессоров и 14 доцентов.

Кафедра имеет 4 компьютерных класса, объединяющих более 70 компьютеров в локальную вычислительную сеть кафедры и обеспечивающих доступ студентов ко всем информационным ресурсам кафедры и выход в Интернет. Кроме того, на кафедре имеются учебные и научно-исследовательские лаборатории по вычислительной технике, в которых работают студенты кафедры.

Чему мы учим

Традиционно на кафедре вычислительной техники Университета ИТМО основной упор в подготовке специалистов делается на фундаментальную базовую подготовку в рамках общепрофессиональных и специальных дисциплин, охватывающих наиболее важные разделы вычислительной техники: функциональная схемотехника и микропроцессорная техника, алгоритмизация и программирование, информационные системы и базы данных, мультимедиа-технологии, вычислительные сети и средства телекоммуникации, защита информации и информационная безопасность. В то же время, кафедра предоставляет студентам старших курсов возможность специализироваться в более узких профессиональных областях в соответствии с их интересами.

Специализации на выбор

Кафедра вычислительной техники Университета ИТМО ведёт подготовку специалистов высшей квалификации в соответствии с Государственными образовательными стандартами 3-го поколения (ГОС-3) по двум направлениям:

09.04.01 «Информатика и вычислительная техника» (профиль подготовки «Вычислительные машины, комплексы, системы и сети»);

09.04.04 «Программная инженерия» (профиль подготовки «Разработка программно-информационных систем»);

с присвоением степени (квалификации) бакалавр (срок обучения – 4 года).

Прием абитуриентов на указанные направления подготовки бакалавров осуществляется в соответствии с общими [Правилами приема в Университет ИТМО](#).

Студенты, успешно завершившие обучение и получившие дипломбакалавра, могут продолжить обучение в магистратуре кафедры (срок обучения – 2 года) по следующим магистерским программам:

- [«Безопасность вычислительных систем и сетей»](#) – руководитель д.т.н. профессор [Щеглов Андрей Юрьевич](#);
- [«Вычислительные системы и сети»](#) - руководитель д.т.н. профессор [Алиев ТауфикИзмайлович](#);
- [«Информационно-вычислительные системы»](#)- руководитель д.т.н. профессор [Алиев ТауфикИзмайлович](#);
- [«Интеллектуальные информационные системы»](#) – руководитель д.т.н. профессор [Тропченко Александр Ювенальевич](#);
- [«Проектирование встроенных вычислительных систем»](#) - руководитель д.т.н. профессор [Платунов Алексей Евгеньевич](#);
- [«Системотехника интегральных вычислителей. Системы на кристалле»](#) – руководитель д.т.н. профессор [Платунов Алексей Евгеньевич](#);
- [«Сетевые встроенные системы»](#) - руководитель д.т.н. профессор [Платунов Алексей Евгеньевич](#);
- [«Технологии компьютерной визуализации»](#) (совместно с базовой кафедрой Института Прикладной математики им. М.В. Келдыша) – руководитель д.т.н. профессор [Палташев Тимур Турсунович](#).

В магистратуру на конкурсной основе принимаются выпускники других вузов, имеющие диплом бакалавра.

На кафедре вычислительной техники Университета ИТМО в рамках аспирантуры и докторантуры осуществляется подготовка научных кадров по следующим специальностям:

- 05.13.05 – Элементы и устройства вычислительной техники и систем управления (технические науки);
- 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей (технические науки);
- 05.13.12 – Системы автоматизации проектирования (приборостроение) (технические науки);
- 05.13.15 – Вычислительные машины, комплексы и компьютерные сети (технические науки);
- 05.13.17 – Теоретические основы информатики (технические науки);
- 05.13.18 – Математическое моделирование, численные методы и комплексы программ (технические науки);
- 05.13.19 – Методы и системы защиты информации, информационная безопасность (технические науки).

**Тропченко Александр Ювенальевич
Тропченко Андрей Александрович**

Методы вторичной обработки и распознавания изображений

Учебное пособие

В авторской редакции

Редакционно-издательский отдел Университет ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати ...

Заказ №

Тираж

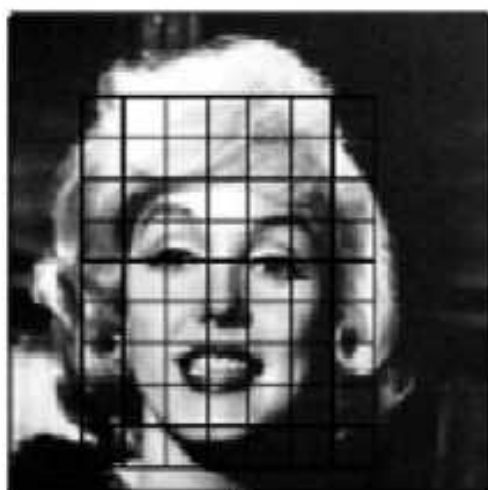
Отпечатано на ризографе

Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49

А. Ю. Тропченко, А.А. Тропченко

**МЕТОДЫ ВТОРИЧНОЙ ОБРАБОТКИ
И РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ**

Учебное пособие



(a)



(b)

**Санкт-Петербург
2015**