

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

**УНИВЕРСИТЕТ ИТМО**

**А.С. Потапов**

**СИСТЕМЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ**

**Учебное пособие**

 **УНИВЕРСИТЕТ ИТМО**

**Санкт-Петербург**

**2016**

Потапов А.С. Системы компьютерного зрения. Учебное пособие. – СПб: Университет ИТМО, 2016. – 161 с.

Рассмотрены сведения из области анализа изображений, необходимые для изучения систем компьютерного зрения. В частности, разобрано ключевое понятие представления изображений и описаны основные уровни представлений изображений, использование которых разобрано на методах сопоставления изображений. Далее рассмотрены основные задачи компьютерного зрения, включающие восстановление трехмерной информации по изображениям, восстановление карт отражательной способности и анализ движения.

Пособие подготовлено в рамках выполнения государственного задания.

Пособие адресовано студентам, обучающимся по направлениям 12.04.03 – "Фотоника и оптоинформатика", а также аспирантам и специалистам, интересующимся проблематикой компьютерной фотоники.

Рекомендовано к печати Ученым советом мегафакультета фотоники, протокол № 03 от 16 ноября 2016 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, 2016

© Потапов Алексей Сергеевич, 2016

## Оглавление

1. Развитие подходов к построению систем компьютерного зрения.....	4
1.1. Проблематика компьютерного зрения.....	4
1.2. Основные сведения о строении зрительной системы приматов.....	10
1.3. Представления изображений.....	18
2. Сопоставление изображений.....	37
2.1. Геометрические преобразования изображений.....	37
2.2. Корреляционные методы сопоставления изображений.....	48
2.3. Контурные и структурные методы сопоставления изображений.....	63
3. Восстановление трехмерной структуры сцены.....	86
3.1. Проблема стереозрения.....	86
3.2. Сопоставление изображений трехмерных сцен.....	98
3.3. Методы восстановления трехмерной формы и дальности.....	114
4. Восстановление физических характеристик объектов и понимание изображений.....	124
4.1. Восстановление отражательной способности.....	124
4.2. Анализ движения.....	136
4.3. Понимание изображений.....	145
Кафедра компьютерной фотоники и видеоинформатики.....	158

# 1. Развитие подходов к построению систем компьютерного зрения

## 1.1. Проблематика компьютерного зрения

Широко известно, что с помощью зрения человек получает наибольший объем информации об окружающих предметах, причем без непосредственного взаимодействия с ними. Однако зрение дается людям настолько естественно, что мало кто задумывается, насколько процесс зрения является сложным. Обычно полагают, будто глаз – это своего рода камера, которая строит изображение окружающего мира, и одного этого изображения уже достаточно для осуществления процесса зрения. Но если глаз просто строит изображение, кто же это изображение дальше рассматривает и как этот процесс все-таки осуществляется?

Действительная сложность зрения была оценена, только когда начались попытки промоделировать его на компьютере. Перспектива снабдить технические системы зрением была очень привлекательной, поскольку это позволило бы им ориентироваться в окружающей обстановке так же легко, как и людям. Однако первые же эксперименты выявили огромные трудности, связанные с наделением машины способностью видеть. Оказалось, что изображения одного и того же объекта, полученные в естественных условиях, настолько не похожи друг на друга, что их очень сложно распознать. Не так просто оказалось осуществлять и навигацию в пространстве по изображениям, поскольку в явном виде плоские изображения не содержат информацию о трехмерных характеристиках окружающих объектов и расстояниях до них. Попытки решить данные проблемы привели к возникновению новой области знаний – компьютерного зрения.

Компьютерное зрение – достаточно молодая (ее возникновение можно отнести к 1950-м гг.) и быстро развивающаяся область научных и прикладных исследований, основной целью которых является построение систем, способных «видеть», то есть извлекать из изображений информацию об объектах внешнего мира, полезную для дальнейшего использования в рамках какого-либо приложения. При этом обработка видеoinформации осуществляется на универсальных или специализированных компьютерах.

Изображение, поступающее на вход системе компьютерного зрения, не обязательно является тем цветным изображением, которое воспринимается человеком. В качестве изображения может выступать любой пространственно-организованный массив измерений некоторых физических величин. В частности, если изображения формируются как результаты измерения интенсивности электромагнитного излучения в различных направлениях, то они могут быть разделены на классы:

- оптические изображения;
- инфракрасные изображения;
- ультрафиолетовые изображения;

- рентгеновские изображения;
- изображения в радиодиапазоне.

Кроме того, в точках изображений могут быть записаны результаты измерений и других физических величин (дальности, плотности и т.д.).

Помимо этого изображения могут различаться объемом информации, хранящемся в каждом пикселе (и, в частности, числом каналов, то есть размерностью вектора измеряемых величин). Здесь можно выделить

- бинарные изображения;
- полутоновые изображения;
- цветные изображения;
- мульти- и гиперспектральные изображения.

Кроме того, и сами изображения могут быть не только двумерные, но и объемные (что особенно характерно для томограмм). Дополнительно можно выделить видеоряды, которые являются трехмерными (в случае двумерных кадров) массивами измерений, но обладают определенной спецификой.

По способу освещения системы компьютерного зрения могут быть разделены на *пассивные* – работающие с изображениями, полученными в естественных условиях освещения, и *активные* – использующие некоторый генератор излучения (например, радиолокационные системы, лазерные дальномеры, системы со структурированной подсветкой и т.д.). Другого типа активность системы компьютерного зрения может заключаться в автоматическом изменении ракурса съемки, фокусного расстояния или каких-либо других параметров объектива.

Таким образом, любые пространственно-организованные массивы данных можно трактовать как изображения, а значит, изображения как источник информации встречаются очень часто. В качестве наиболее часто встречаемых классов изображений, принадлежащих различным предметным областям, можно назвать

- изображения, полученные внутри помещений;
- (наземные) изображения, полученные вне помещений;
- аэрокосмические изображения;
- биомедицинские изображения;
- микроскопические изображения.

Задачи компьютерной обработки и анализа изображений открывают широкие перспективы автоматизации многих сфер человеческой деятельности. Наиболее очевидным приложением является очувствление роботов. Действительно, сложно себе представить робота, способного адаптироваться к каким-либо изменениям ситуации без органов чувств. Адаптивные роботы имеют разнообразные применения в производстве при сборке изделий, в быту (помогающие в хозяйстве или больным людям, например, роботы-поводыри, или выступающие в роли домашних животных), при исследовании других планет или опасных для человека мест на Земле. В качестве других приложений можно назвать неразрушающий контроль качества выпускаемой продукции, навигацию беспилотных

летательных аппаратов или автомобилей, системы безопасности и экологического мониторинга, биометрии, умные фото и видеокамеры, системы поиска изображений в базах данных (как в домашних фотоальбомах или интернете, так и в геоинформационных системах), обнаружение, распознавание и сопровождение целей и так далее.

Следует отметить, что разрабатываемые в настоящее время системы компьютерного зрения предназначены для решения конкретных задач, обычно ориентированы на какой-то один из перечисленных выше типов изображений и работают в ограниченных предметных областях.

Построение системы компьютерного зрения общего назначения является неразрешимой на текущий момент задачей и может рассматриваться лишь как некая конечная цель исследований в данной области. В этом смысле компьютерное зрение похоже на область искусственного интеллекта, в которой также моделируются лишь отдельные аспекты интеллектуальной деятельности. Если искусственный интеллект трактовать как область, целью которой является снабжение компьютеров такими возможностями по обработке информации, которые имеют живые организмы, то в качестве одной из подцелей является наделение машин возможностями по обработке сенсорной информации. В этом смысле компьютерное зрение может рассматриваться как раздел искусственного интеллекта. Более того, вполне вероятно, что построение систем компьютерного зрения общего назначения невозможно в полной мере без решения проблемы искусственного интеллекта.

В то же время, системы компьютерного зрения существенно опираются на специфику обрабатываемой информации, в связи с чем данная область имеет собственный предмет и методы исследования, то есть является отдельным направлением науки. При этом она также пересекается с такими областями, как компьютерная графика, обработка изображений, психология восприятия и т.д.

Существует ряд терминов, так или иначе относящихся к выполнению операций над изображениями с помощью компьютеров:

- компьютерное зрение (а также техническое зрение, машинное зрение, зрение роботов);
- обработка изображений;
- анализ изображений;
- интерпретация изображения;
- понимание изображений;
- иконика;
- распознавание образов;
- компьютерная графика;
- когнитивная графика.

В то время как компьютерное, машинное и техническое зрение можно считать почти эквивалентными терминами, хотя и имеющим свой смысловой оттенок (зрение роботов можно трактовать как несколько более узкое), все

остальные термины обозначают достаточно близкие, но все же различные понятия.

Все они связаны с преобразованием изобразительной информации, однако отличаются тем, что именно подается на вход, а что – на выход. В случае обработки изображений как входной, так и выходной информацией является изображение (хотя иногда обработку изображений понимают более широко – как любое преобразование изображений, в том числе и к неизобразительной форме). Типичными задачами обработки изображений являются задачи улучшения контраста, подавления шумов, подчеркивания границ, реставрации изображений и т.д.

Под интерпретацией изображений обычно понимается преобразование изображения в некоторую символьную (неизобразительную) форму с извлечением некоторой значимой информации. К интерпретации изображений можно отнести задачи текстурного анализа, построения структурных описаний изображений (в терминах областей, контуров, геометрических элементов) или выделения «точек интереса». Анализ изображений является более широким понятием и, помимо задач интерпретации изображений, также включает такие задачи, как нахождение взаимного пространственного преобразования двух изображений одной и той же сцены, снятой с разных ракурсов, выявление произошедших изменений по серии снимков и т.д. Под пониманием изображений также подразумевает их преобразование к неизобразительной форме, однако здесь на выходе формируются семантические описания изображений (то есть описания в терминах объектов и взаимосвязей между ними). Иногда термин «интерпретация изображений» используется в том же смысле, что и термин «понимание изображений».

Под иконикой понимается общая наука об изображениях. Сюда могут относиться как, например, вопросы улучшения изображений (в смысле их визуального восприятия человеком), так и вопросы автоматического анализа изображений. В то же время, здесь основной акцент делается на исследование природы изображений как таковых, а компьютерные методы их преобразования выступают лишь в качестве приложений.

Компьютерная и когнитивная графика связаны с обратным преобразованием: от символьных представлений к изображениям (хотя как исходные данные, так и результирующие изображения в этих областях существенно различаются).

Распознавание образов представляет собой отдельную научную дисциплину, с изображениями связанную лишь косвенно. В ней под образом понимается произвольное описание объекта. При этом если при распознавании изображений используются методы распознавания образов, то только после извлечения из изображений некоторых инвариантных признаков.

Компьютерное зрение, составляющее предмет изучения в рамках настоящего курса, теснее всего связано с анализом и пониманием изображений, но имеет свою специфику. Отличительной особенностью задач

компьютерного зрения является то, что в них на основе изображений требуется восстановить информацию об объектах, имеющую физический смысл, в связи с чем оказывается необходимым использовать детальную информацию о процессах распространения излучения и формирования изображений оптическими системами. Напротив, в методах анализа изображений могут строиться описания, никак не связанные с какими-либо физическими величинами, а сами методы могут быть одинаково применимы для изображений разной природы. Наиболее типичными задачами компьютерного зрения являются задачи

- построения карт отражательной способности по видимой яркости поверхностей;
- определения взаимных перемещений объектов сцены и приемника излучения;
- восстановления трехмерной информации об объектах методами
  - стереозрения,
  - структурированной подсветки,
  - фотометрического стерео,
  - формы из затенения,
  - формы из движения,
  - глубины из дефокусировки;
- распознавания объектов.

Напротив, такая задача, как генерация оптимальной камуфляжной раскраски по изображению местности, может быть отнесена к области анализа изображений (хотя включает элемент компьютерной графики), но вряд ли – к компьютерному зрению.

В чем же причина сложности задач компьютерного зрения и анализа изображений? Как уже отмечалось, изображения обладают большой изменчивостью. Смена ракурса съемки, условий освещения, взаимные перемещения объектов и их частей с частичным загоранием приводят к сложным преобразованиям в яркостях пикселей при постоянстве содержания изображения, что называется *изменчивостью изображений*. В результате, отдельные точки изображения и даже весьма большие области могут быть недостаточными для решения задач зрения.

К примеру, цвет конкретного объекта существенно меняется в зависимости от освещения. Белый лист при тусклом освещении будет менее ярким, чем темно-серый лист при ярком освещении. Тем не менее, человек правильно определяет цвета независимо от освещения, что называется *константностью восприятия*.

Освещение часто является очень сложным: может быть несколько источников света (причем, не только точечные, но и протяженные), какие-то объекты могут находиться в тени, а какие-то – на свету. Все становится еще сложнее, когда возникает вопрос о восстановлении трехмерной информации, которая в явном виде в изображении не содержится. Распознавание объектов по изображениям и семантическое описание сцен является наиболее сложной задачей, не только в связи с указанными причинами изменчивости, но и

большим разнообразием объектов и их сложной внутренней структурой и пространственными отношениями.

Представьте, что вы смотрите на аэрокосмическую фотографию (рис. 1.1а), и вас просят найти на ней корабль. Вы никогда не видели корабль на фотографии такого типа и с такого ракурса, но у вас не будет существенных проблем решить эту задачу (см. рис. 1.1б). В то же время по небольшому фрагменту изображения без дополнительной информации вы вряд ли сможете распознать объект, представленный на рис. 1.1в.

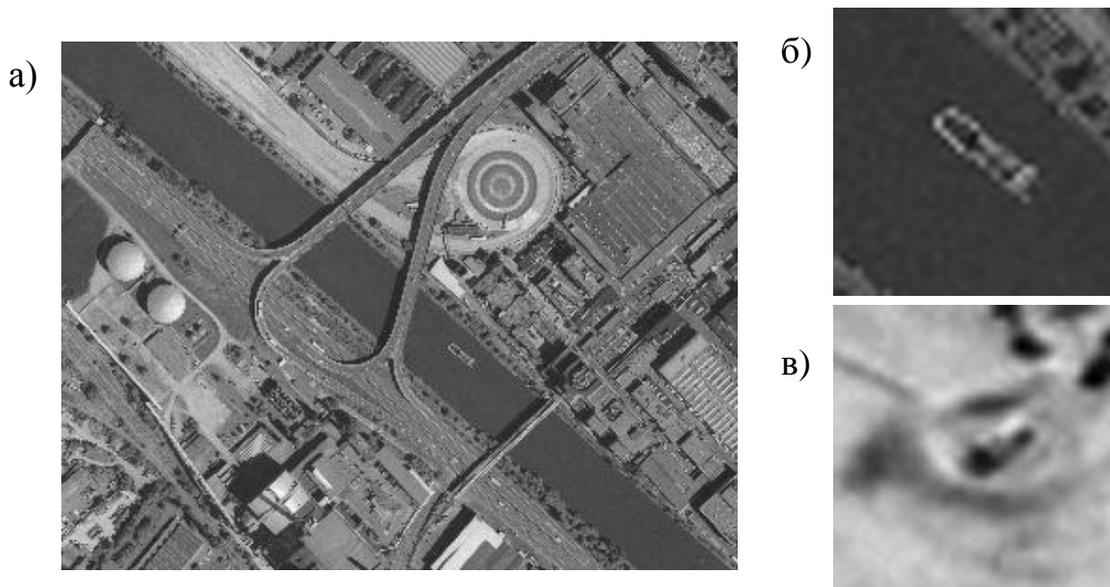


Рис. 1.1: (а) аэрокосмический снимок, на котором присутствует корабль; (б) увеличенный фрагмент с кораблем (в): увеличенный фрагмент изображения, содержащий объект, сложный для опознавания без контекста

Таким образом, для изображений свойственна не только большая изменчивость, но также и *неоднозначность* их интерпретации в условиях высокой *априорной неопределенности* (означающей, что возможное содержание изображения заранее неизвестно, и на нем могут присутствовать самые разные объекты). Как видно из рис. 1.1а, при распознавании широко используется контекстная информация, то есть восприятие обладает таким свойством, как *целостность* (интерпретация изображения не сводится лишь к интерпретации его фрагментов).

Как же достигается константность восприятия и преодолевается неоднозначность (или реализуется целостность восприятия)? Поскольку первые эксперименты в области компьютерного зрения по опознаванию объектов путем непосредственного сравнения их изображений с эталонными изображениями, не позволили получить универсальных решений, естественным было обратиться к нейрофизиологическим и психофизическим исследованиям зрительной системы животных и человека, которые указанные проблемы умеют решать.

Хотя, возможно, заимствование у биологических систем зрительного восприятия было и не определяющим в развитии компьютерного зрения, оно все же оказало определенное влияние и в целом соответствует логике

развития данной области (и, к тому же, компьютерная парадигма в психологии и нейрофизиологии дала определенный толчок к развитию этих наук).

### Вопросы и упражнения

1. В чем различие между компьютерным зрением, анализом изображений, компьютерной графикой и обработкой изображений? Какие еще дисциплины, связанные с изображениями, можно назвать?
2. Что такое изображение? На какие классы могут быть разделены изображения?
3. Каковы типичные задачи и приложения компьютерного зрения?
4. В чем заключаются основные трудности, возникающие при построении систем компьютерного зрения?
5. Что такое константность восприятия?

### 1.2. Основные сведения о строении зрительной системы приматов

Одной из первых попыток использования бионического подхода (то есть подхода, в рамках которого происходит применение идей, почерпнутых у биологических систем) в компьютерном зрении стало использование искусственных нейронных сетей в целях распознавания объектов по их изображениям. Более того, развитие самого нейросетевого подхода во многом связано именно с попытками решения проблемы зрительного восприятия.

Формальная модель нейрона, как порогового сумматора (см. рис. 1.2), была предложена Мак-Каллоком и Питсом еще в 1943 году.

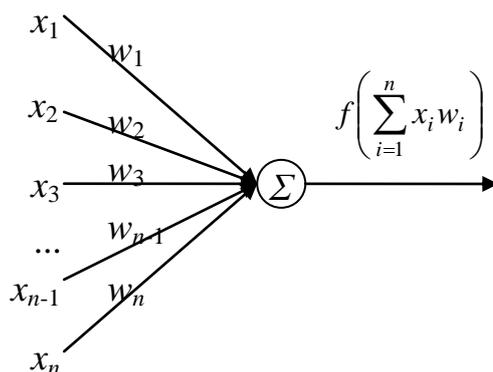


Рис. 1.2. Модель формального нейрона как порогового сумматора:  $x_i$  – входные сигналы от других нейронов;  $w_i$  – веса связей;  $f$  – нелинейная активационная функция

Одной из первых искусственных нейронных сетей (ИНС) с определенной архитектурой и способом назначения весов связей между нейронами стала предложенная Ф. Розенблаттом ИНС, названная им *перцептроном* (от слова *perception* – восприятие).

Архитектура перцептрона представлена на рис. 1.3. На входной слой нейронов, названный исходно сетчаткой и представлявший собой матрицу, подавалось изображение, а активность нейронов выходного слоя должна была говорить о том, объект какого типа присутствует на изображении.

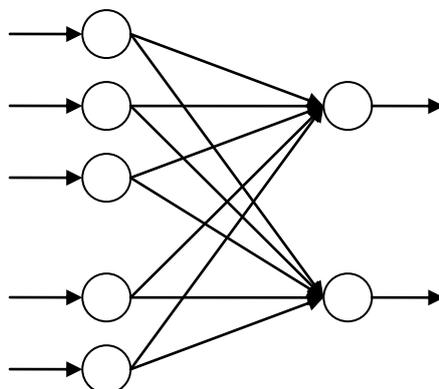


Рис. 1.3. Архитектура однослойного перцептрона

Перцептрон предназначался для моделирования (в самых общих чертах) работы сетчатки, которая, как к тому времени было известно, содержит несколько слоев нейронов. Поскольку в то время было известно крайне мало, как об устройстве зрительной системы, так и о проблеме зрения вообще, у исследователей была надежда на то, что использование самой по себе нейросетевой архитектуры позволит построить универсальную систему распознавания.

Однако эти ожидания не оправдались (проблему не решило и добавление промежуточных слоев нейронов). Перцептрон оказался способным к распознаванию лишь при очень сильных ограничениях. Помимо ограничения на число классов объектов, распознаванию которых может быть обучен перцептрон, данная нейронная сеть не способна распознавать объекты, изображения которых сдвинуты и повернуты относительно эталонных изображений обучающей выборки. В связи с этим, перцептрон стал использоваться для распознавания объектов не по исходным изображениям, а по некоторым признакам объектов, инвариантным к геометрическим преобразованиям. Это использование перцептрона было более успешным на практике, однако оно оставляло вопрос о том, как именно следует строить подобные системы признаков (и, кроме того, перцептрон при этом переставал быть моделью сетчатки).

Дальнейшие более детальные исследования сетчатки показали, что в ней выполняется лишь предварительная обработка изображения. Кратко рассмотрим строение (рис. 1.4) и работу сетчатки.

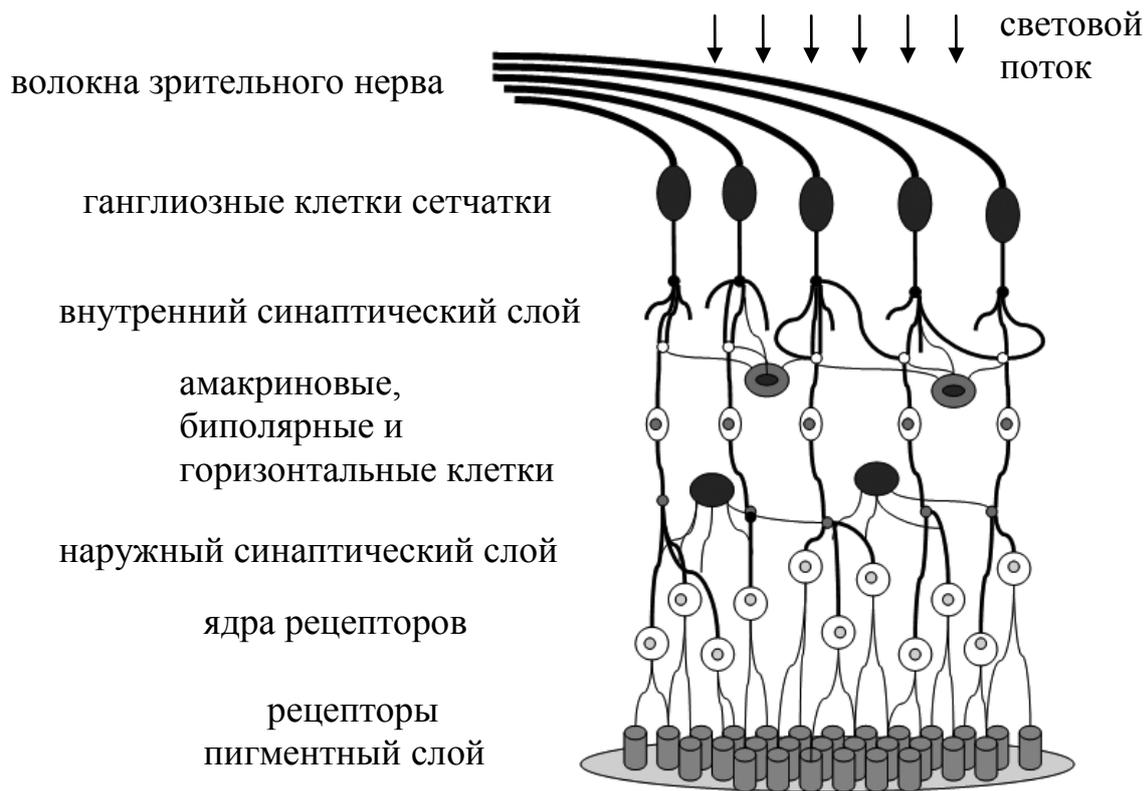


Рис. 1.4. Структура сетчатки

Сетчатка преобразует электромагнитное излучение в нервные сигналы благодаря слою светочувствительных рецепторов (палочек и колбочек). Этот слой, как ни странно, расположен дальше всего от поверхности глаза, но, с точки зрения обработки зрительной информации, является первым. Палочки обеспечивают способность видеть при слабом освещении, в то время как колбочки обеспечивают цветное зрение в условиях хорошего освещения. В каждом глазу имеется в сумме порядка 125 млн. палочек и колбочек, неравномерно распределенных по сетчатке (в центральной ямке, концентрация колбочек максимальна, что обеспечивает высокую остроту зрения).

Вместе со слоем рецепторов сетчатка содержит три крупных слоя нервных клеток, разделенных двумя слоями, образованными дендритами и аксонами этих клеток (толщина сетчатки составляет около четверти миллиметра).

Передний слой сетчатки состоит из *ганглиозных* клеток. Аксоны этих клеток проходят по поверхности сетчатки к слепому пятну, в котором они собираются в зрительный нерв, соединяющий глаз и мозг. Интересно отметить, что ганглиозных клеток, передающих зрительную информацию в мозг из сетчатки, всего около 1 миллиона, то есть примерно в 125 раз меньше, чем число рецепторов.

В среднем слое имеются три типа нейронов: биполярные, горизонтальные и амакриновые. Биполярные клетки получают входные сигналы от сравнительно небольшого числа рецепторов и в большинстве своем направляют выходные сигналы непосредственно ганглиозным

клеткам. Биполярные клетки могут иметь входы не только от рецепторов, но и от горизонтальных клеток, которые соединены с большим числом рецепторов длинными связями, идущими вдоль слоя сетчатки. Как горизонтальные клетки связывают рецепторы и биполярные клетки, так и амакриновые клетки связывают биполярные клетки с ганглиозными (в дополнение к прямым связям между ними).

Для исследования процессов переработки зрительной информации, происходящей в сетчатке, нейрофизиологами изучались отклики ганглиозных клеток при предъявлении различных стимулов в зрительном поле. В частности, Куффлер проецировал на сетчатку маленькое пятнышко света и, перемещая его, отыскивал на сетчатке области, освещение которых увеличивало или подавляло активность конкретной ганглиозной клетки, к которой был подключен электрод. Такая область (то есть область рецепторов, от которых доходят сигналы до некоторого нейрона) называется *рецептивным полем* соответствующего нейрона.

Выяснилось, что рецептивные поля ганглиозных клеток круглые и состоят из центра и периферии (см. рис. 1.5), причем эти клетки бывают двух типов: у клеток с так называемой *on*-реакцией засветка центра рецептивного поля приводит к повышению активности клетки, а засветка периферии – к подавлению ее активности, в то время как у клеток с *off*-реакцией все в точности наоборот.

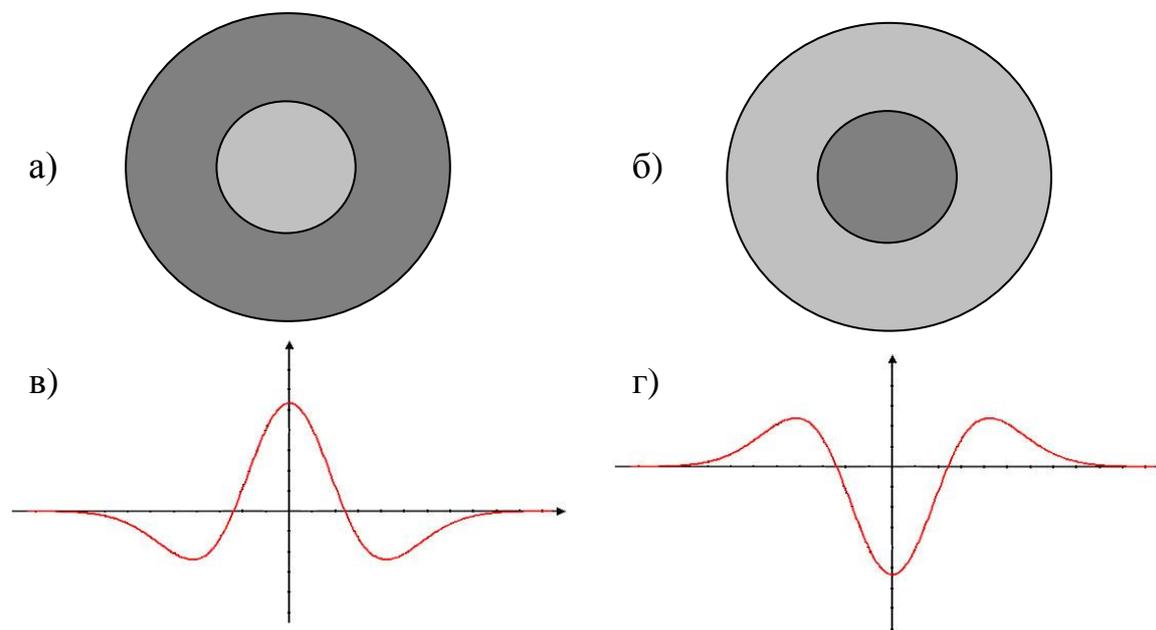


Рис. 1.5. Структура рецептивных полей ганглиозных клеток (темным показаны области торможения, светлым – возбуждения): (а) случай *on*-реакции; (б) случай *off*-реакции; (в) и (г) – сечения карт рецептивных полей в случае *on*- и *off*-реакций, аппроксимированных как разности гауссиан

Таким образом, если в центр рецептивного поля клетки с *on*-реакцией проецировать пятнышко света, то при увеличении размеров этого пятнышка активность клетки будет возрастать, пока пятно не заполнит весь центр; когда же пятно начнет заполнять периферию, то активность клетки будет

уменьшаться (при равномерной засветке всего рецептивного поля ганглиозная клетка практически не будет проявлять активности).

Вполне возможно, что центр рецептивного поля ганглиозной клетки реализуется с помощью прямого пути распространения сигнала от рецепторов через биполярные клетки к ганглиозным клеткам, ведь биполяры получают входы от небольшого числа рецепторов, а ганглиозные клетки – от небольшого числа биполяров. Тогда периферия рецептивного поля ганглиозной клетки реализуется через не прямой путь, включающий горизонтальные и амакриновые клетки, имеющие более обширные связи.

Стоит отметить, что рецептивные поля "соседних" ганглиозных клеток существенно перекрываются, кроме того, размеры рецептивных полей сильно варьируются. Интерпретация действия ганглиозных клеток (совместно с клетками среднего слоя сетчатки) может быть различной. С одной стороны, можно полагать, что происходит локальное вычитание яркости фона для получения относительной яркости объектов в целях достижения константности восприятия по отношению к условиям освещения. С другой стороны, считается, что форма рецептивных полей ганглиозных клеток неплохо описывается разностью двух гауссиан с разной дисперсией (рис. 1.5в и 1.5г), что является аппроксимацией оператора Лапласа. В таком случае, действие ганглиозных клеток можно трактовать как пространственное дифференцирование изображения с выделением на нем контуров.

Рассмотрим кратко дальнейшую структуру зрительного пути (рис. 1.6).

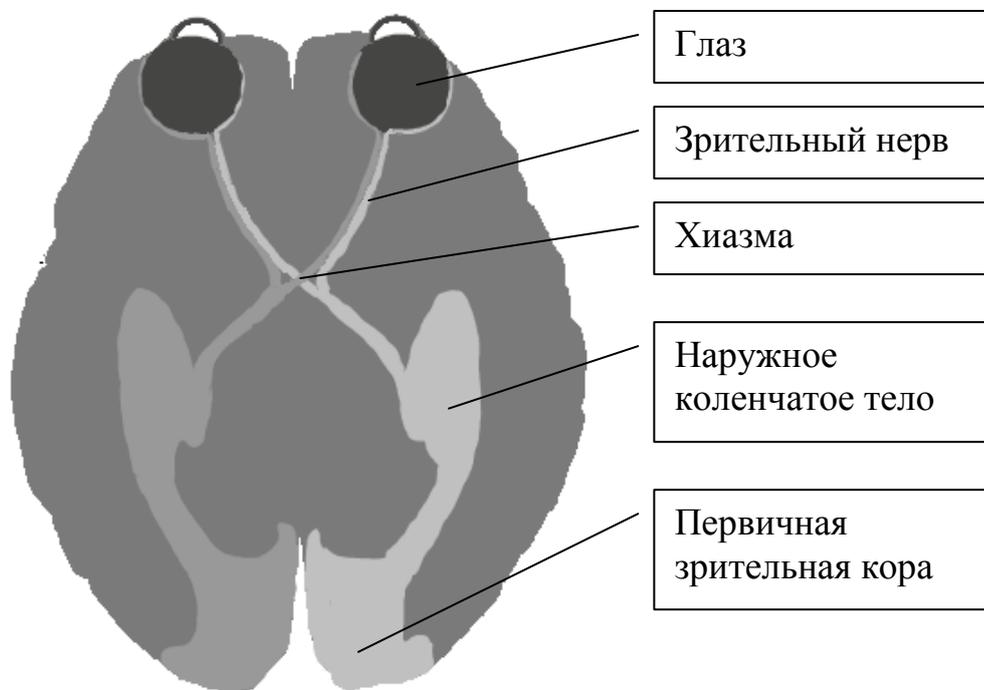


Рис. 1.6. Начальная структура зрительного пути

Аксоны ганглиозных клеток сетчатки (зрительный нерв) входят в наружное коленчатое тело (НКТ), а аксоны клеток НКТ идут далее к

первичной зрительной коре, также называемой стриарной корой или полем 17.

При этом связи между сетчаткой, НКТ и стриарной корой имеют топографическую организацию, то есть сохраняют ту же упорядоченность в пространстве, что и у точек изображения на сетчатке. В частности, при разрушении какой-то области стриарной коры наступает такая же «локальная слепота», как и при разрушении некоторой области на сетчатке. Однако это не означает простого взаимнооднозначного соответствия, поскольку при этом производится сложная обработка информации.

Зрительные нервы, идущие от двух глаз, пересекаются в хиазме (внешний вид которой напоминает букву  $\chi$ ), где расщепляются на две части, идущих к разным полушариям мозга. Таким образом, от каждого глаза информация передается в оба полушария. Общеизвестно, что за управление правой рукой отвечает левое полушарие и наоборот. Можно было бы ожидать, что и здесь информация от правого глаза должна передаваться в левое полушарие. Однако в данном случае в левое НКТ поступает информация не от правого глаза, а от правой половины зрительного поля обоих глаз (и, наоборот, в правое НКТ поступает информация от левой половины зрительного поля обоих глаз).

Получение информации от двух глаз позволяет работать механизмам стереозрения. Перерезка хиазмы ведет к так называемой стереослепоте, однако не полной, а только для тех точек пространства, для которых диспаратность (различие углов, под которыми эти точки видны для двух глаз) достаточно мала. Это вполне естественно, поскольку именно эти точки попадают в одинаковую часть зрительного поля для обоих глаз. Если же значение диспаратности большое (в частности, для близких объектов), то информация от этих точек попадет от правого и левого глаза в разные полушария. Работа механизмов стереозрения для этих точек возможна благодаря передаче информации между полушариями через *мозолистое тело* – самый крупный пучок волокон в нервной системе (состоящий из примерно 200 млн. аксонов), соединяющий два полушария (по нему передается далеко не только зрительная информация). При перерезке мозолистого тела наступает стереослепота для точек пространства с высокой диспаратностью.

Не все волокна зрительного нерва, пройдя хиазму, идут в НКТ. Небольшая их часть передает информацию специфическим группам нейронов, имеющих отношения к управлению движением глаз.

Каждое из НКТ содержит примерно 1.5 миллиона нейронов, рецептивные поля которых весьма сходны с рецептивными полями ганглиозных клеток сетчатки. Однако, помимо входов от ганглиозных клеток, нейроны НКТ имеют входы от соседних клеток НКТ, обратные входы от тех корковых зон, в которые передается информация от НКТ, а также входы от ретикулярной формации, связанной с управлением вниманием. Все это может иметь отношение к тонким механизмам разрешения неоднозначности в процессе зрительного анализа, но в целом в НКТ зрительная информация преобразуется несущественно (даже входы от

двух глаз в НКТ разделены по слоям, и возможности стереозрения еще не реализуются).

Напротив, нейроны стриарной коры обладают принципиально новыми свойствами. В то время как ганглиозные клетки и клетки НКТ обладают круговой симметрией, большая (более 70%) часть нейронов зрительной коры обладает свойством дирекционной избирательности (чувствительности к ориентации стимула).

Стриарная кора имеет несколько слоев, и, естественно, с каждым следующим слоем реакции нейронов усложняются. Первые (в порядке обработки информации в зрительной системе) нейроны с дирекционной избирательностью были названы *простыми клетками*. Эти нейроны бывают разных типов, чувствительных к трем видам стимула (рис. 1.7).

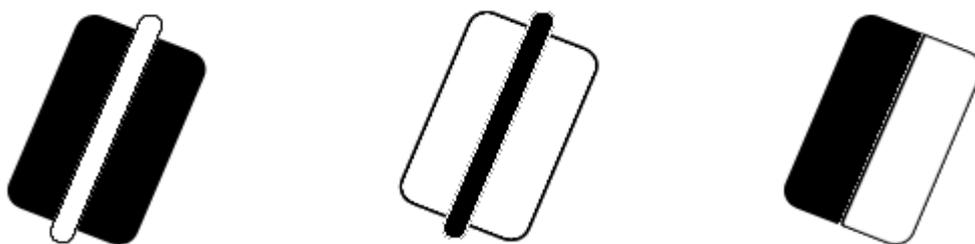


Рис. 1.7. Три типа стимулов (светлая полоса на темном фоне, темная полоса на светлом фоне и граница), к которым могут быть чувствительны клетки стриарной коры

Ключевой характеристикой является ориентация стимула. Как правило, простые клетки наиболее сильно реагируют на линию или край одной ориентации, а при отклонении угла ориентации на 10-20 градусов реакция клетки существенно уменьшается. Кроме того, наиболее сильную реакцию простой клетки вызывает стимул при определенном положении, при отклонении от которого реакция сильно снижается.

Следующему уровню анализа соответствуют *сложные клетки*, получающие входы от простых клеток. Эти клетки также чувствительны к стимулам трех указанных видов и обладают дирекционной избирательностью. Однако в отличие от простых клеток они реагируют на соответствующий стимул независимо от того, в каком месте рецептивного поля он появляется. Помимо этого предпочтение часто отдается движущимся в определенном направлении стимулам (реакция же простой клетки на движущийся стимул проявляется только в тот момент, когда он проходит центр ее рецептивного поля). Также в зрительной коре присутствуют бинокулярные клетки (то есть клетки, получающие информацию от обоих глаз).

Интересно отметить, что стриарная кора содержит около 200 миллионов нейронов (при этом они получают входы всего от двух миллионов ганглиозных клеток), каждый из которых чувствителен к собственному "элементарному" стимулу. Однако на этом обработка зрительной

информации не заканчивается, ведь помимо первичной зрительной коры существует порядка полутора десятков высших корковых зон, вовлеченных в зрительное восприятие (они составляют около 15% площади всей коры головного мозга). Каждая из этих зон передает информацию в несколько последующих зон. Однако о процессах преобразования информации в этих зонах детальных сведений нет. Во второй зрительной зоне, по-видимому, есть три группы клеток, отвечающих за механизмы стереозрения и восприятия цвета и формы. Последующие зоны также отвечают за конкретные аспекты зрительного восприятия с постепенным возрастанием степени абстрактности и общности результатов преобразования исходного изображения.

Все это показывает, что зрение – очень сложный процесс, в котором из изображения постепенно извлекается информация все более высокого уровня абстракции, и, кроме того, информация обрабатывается в нескольких каналах, таких как цвет, форма, текстура, движение и т.д. На начальных уровнях этого анализа, по-видимому, выделяются яркостные перепады, на основе которых осуществляется обнаружение линий и полос.

Сходные результаты показывают и психофизические опыты, в которых определялось, в какие точки на изображениях преимущественно направлен взгляд человека. Как оказалось, в качестве таких информативных точек изображения часто выступают концы линий, их изломы и перекрестия. Подобные объекты можно трактовать как структурные элементы изображения.

Кроме того, психологические исследования зрительного восприятия показывают, что в мозгу происходит группирование объектов зрительного поля с выделением более сложных структурных образований (см. рис. 1.8).

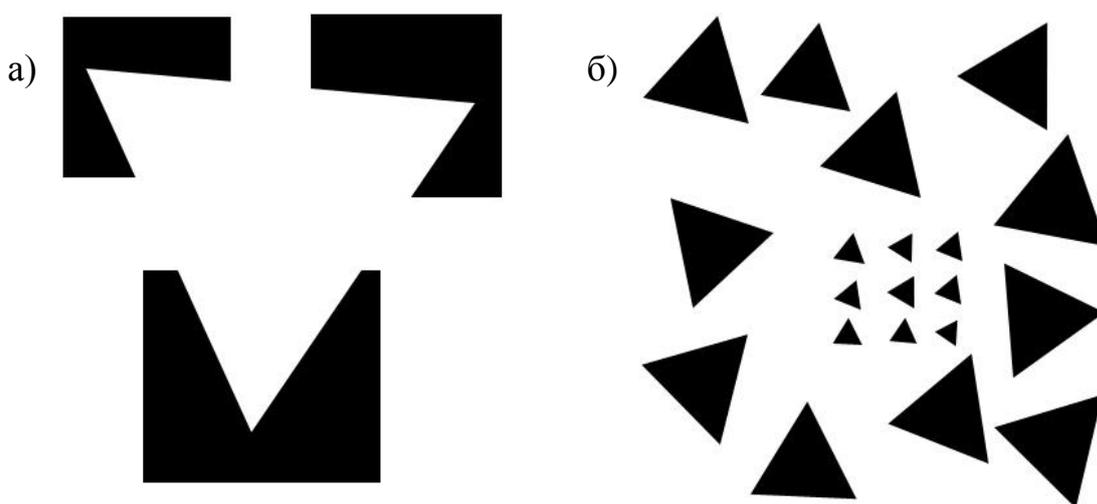


Рис. 1.8. Примеры изображений, иллюстрирующих законы группирования (а) по «хорошему продолжению» и (б) по подобию

Выделяют законы перцептивного группирования по подобию (в группы имеют тенденции объединяться сходные или подобные элементы зрительного поля), близости, «хорошему продолжению» (объединяются

элементы, располагающиеся на некоторой регулярной кривой, например, линии), замкнутости формы и т.д.

### Вопросы и упражнения

1. Чем вызвана неудача применения перцептрона в задачах распознавания изображений?
2. Какие типы светочувствительных рецепторов и нервных клеток есть в сетчатке?
3. В чем основная особенность рецептивных полей ганглиозных клеток?
4. Какие области мозга непосредственно участвуют в обработке зрительной информации?
5. В чем основная особенность рецептивных полей клеток стриарной коры, и чем отличаются простые и сложные клетки?
6. Каких клеток больше: светочувствительных рецепторов, ганглиозных клеток сетчатки или клеток стриарной коры?
7. Какие области мозга участвуют в обеспечении стереоскопического зрения?
8. Как много высших корковых зон в зрительной коре?
9. Охарактеризуйте информативные точки изображений, выявленные в психофизических экспериментах.

### 1.3. Представления изображений

В процессе зрительного восприятия происходит отображение зрительного входа на ранее возникшие модели мира. Существует большой разрыв между изображением и их семантическим содержанием. В области компьютерного зрения постепенно сложилось устойчивое мнение, что в процессе анализа изображений должны строиться некоторые промежуточные описания.

Описание любого объекта, в частности изображения, всегда производится в рамках некоторого *представления*. Под представлением понимается формальная система, содержащая алгоритмы для получения в явном виде описаний объектов заданного класса. Наряду с термином «описание объекта» мы будем использовать термин «модель объекта». Тогда зрение может рассматриваться как процесс преобразования информации, то есть как процесс преобразования описаний изображений из некоторого исходного представления в конечное представление через промежуточные (что, как указывалось ранее, обосновано также нейрофизиологическими и психофизическими данными). Исходное представление определяется способом получения изображений и обычно является фиксированным.

Любой алгоритм компьютерного зрения или анализа изображений использует некоторое внутреннее представление изображений для решения

конкретной задачи зрения. Таким образом, понятие представления оказывается одним из ключевых.

В области компьютерного зрения и анализа изображений разработано множество представлений изображений. В общих чертах их можно разделить по трем уровням абстракции: нижний (или пиксельный) уровень, средний (или промежуточный символьный) уровень и верхний (или семантический) уровень. Каждый из этих уровней включает большое разнообразие различных представлений и может быть дальше уточнен. Такое деление также не вполне строгое: некоторые представления могут быть отнесены к разным уровням. В частности, здесь пиксельный уровень отделен от математических представлений, а контурные представления отнесены к промежуточным символьным.

### *Представления изображений в виде необработанных данных*

В задачах компьютерного зрения в качестве исходного представления, из которого осуществляется отображение в некоторое конечное представление, обычно выступает представление изображения в виде массива «сырых» данных – набора результатов физических измерений (о них говорилось ранее), выполненных для некоторой сцены. Под сценой обычно понимается набор физических объектов или некоторый фрагмент реального мира.

Мельчайшим элементом изображения при таком начальном представлении является пиксель (от pixel: picture element), содержащий результат единичного измерения данной физической величины. В связи с этим, такие представления также называют представлениями на уровне пикселей.

Пиксели, образующие изображение, с соответствующими им значениями интенсивностей организованы в массив, размерность которого определяется природой данных. Упорядочение пикселей в этом массиве соответствует пространственной организации сцены.

Основной проблемой, связанной с представлениями на пиксельном уровне, является проблема эффективности этих представлений в задачах их хранения и передачи. Это особенно актуально для объемных изображений, для которых существует несколько альтернативных способов представления, таких как, например, трехмерный массив вокселей, стереопара, проекция на Гауссову сферу и т.д.

Описания изображений в рамках представлений пиксельного уровня содержат в себе всю имеющуюся информацию о наблюдаемой сцене, но в форме, неудобной для автоматического анализа. 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 говорят о том, какие операции можно выполнить над изображениями. В связи с этим возникает естественное желание представить изображение как элемент некоторого математического пространства. Часто отображение из пиксельного представления в выбранное математическое пространство является взаимнооднозначным и непосредственно выражается через исходные значения яркостей, а результаты математических операций над изображениями снова представляются в виде изображения. Из-за этого интерпретации изображений в качестве массивов пикселей и в качестве математических объектов часто смешивают, объединяя их в один класс, а сам процесс такой переработки изображений выделяют в отдельный тип задач – обработка изображений. Хотя математические модели изображений можно условно отнести к пиксельному уровню, между ними существуют принципиальные отличия.

Интерпретация изображения в качестве элемента математического пространства позволяет распространить формальные операции, введенные на этом пространстве, и на изображения. Это дает обширный набор строгих внутренне непротиворечивых средств анализа и преобразования изображений. К примеру, если выбранное пространство является метрическим, то появляется возможность формального определения "расстояния" (т.е. степени сходства) между изображениями.

К сожалению, существующие на данный момент строгие математические модели изображений являются достаточно низкоуровневыми и имеют ограниченную область применения.

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

### *Функциональные модели*

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

$$f : G \rightarrow V, G \subseteq R^n, V \subseteq R^m, \quad (1.1)$$

где  $G$  – область определения функции, а  $V$  – область ее значений. Величина  $n$  – размерность изображения (обычно равная двум), а  $m$  – размерность вектора физических величин, измеренных для каждой точки. 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 яркости, фильтрация и др.

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

$$f_2(\mathbf{x}') = f(g(\mathbf{x})), g : G \rightarrow G_2, \quad (1.2)$$

где  $g$  – это функция, ставящая в соответствие каждой точке из области определения  $G$  исходного изображения  $f$  точку в области определения  $G_2$  преобразованного изображения  $f_2$ .

Преобразование яркости (например, изменение яркости или контраста для всего изображения) или преобразования цвета (если  $V$  – это цветовое пространство) может быть представлено следующим образом:

$$f_2(\mathbf{x}) = h(f(\mathbf{x})), h: V \rightarrow V_2, \quad (1.3)$$

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

Фильтрация изображений является более широким понятием, чем два приведенных выше преобразования, так как в общем случае зависит от всего содержимого изображения. Если  $\Phi = \{f \mid f: G \rightarrow V\}$  – это функциональное пространство, к которому принадлежит данное изображение, а  $\Phi_2$  – пространство, к которому принадлежит обработанное изображение, то фильтрация изображения – это произвольное отображение из пространства  $\Phi$  в  $\Phi_2$ :  $T: \Phi \rightarrow \Phi_2$ .

Наиболее часто привлекаемой является линейная фильтрация, которая представима в виде операции свертки с некоторым ядром  $\varphi$ :

$$f_2 = T(f) = f \otimes \varphi, (f \otimes \varphi)(\mathbf{x}) = \int_G f(\mathbf{y})\varphi(\mathbf{x} - \mathbf{y})d\mathbf{y}, \quad (1.4)$$

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

Еще одной важной возможностью, предоставляемой функциональными моделями изображений, является смена базиса в исходном функциональном пространстве  $\Phi$ . Частным, но очень важным, примером является преобразование Фурье:

$$F(\boldsymbol{\omega}) = \int_G f(\mathbf{x})e^{-i(\boldsymbol{\omega}, \mathbf{x})} d\mathbf{x}. \quad (1.5)$$

Функциональные (и другие математические) представления позволяют формально ввести понятие инварианта. Как отмечалось, желание работать с представлениями, инвариантными к некоторому типу преобразований, совершенно естественно проистекает из того факта, что сцена с одним и тем же содержанием может совершенно по-разному (при пиксельном представлении) выглядеть при разных условиях наблюдения.

Поскольку каждое изображение исходно представляется в виде набора точек с соответствующими им значениями яркостей (конечного массива данных), а число функций данного класса, как правило, бесконечно, то необходимо вводить некоторые ограничения, позволяющие из множества всех подходящих функций выбрать лучшую.

Часто это достигается тем, что рассматривается ограниченный класс функций, так что по набору точек, образующих данное изображение, выбор функции однозначен. Однако в ряде случаев удобнее накладывать

интегральные ограничения из некоторых модельных соображений. Эти ограничения часто представляются в виде минимизации некоторого функционала, и задача построения модели изображения превращается в задачу вариационного исчисления.

Наиболее простыми функционалами, подвергающимися минимизации, являются следующие:

$$L_2(f) = \int_G |f(x)|^2 dx, \quad L_1(|\nabla f|) = \int_G |\nabla f(x)| dx, \quad \text{и} \quad L_2(|\nabla f|) = \int_G |\nabla f(x)|^2 dx. \quad (1.6)$$

Последняя норма является наиболее широко используемой (это т.н. регуляризация Тихонова).

Выбор минимизируемого функционала (также называемого мерой сложности изображения) часто осуществляется на основе исследований, посвященных определению статистических свойств естественных изображений и особенностей зрительного восприятия животных и человека. Этот подход также применяется при привлечении статистических моделей изображений: если изображение представляется в виде случайной функции, то выбор ее наиболее вероятной реализации осуществляется посредством минимизации некоторого функционала при соблюдении ограничений, налагаемых исходными данными.

### *Вероятностные модели*

В общем случае вероятностные модели являются более общими, чем функциональные, так как представляют изображения некоторыми случайными функциями, по отношению к которым регулярные функции являются весьма частным подклассом. Однако зачастую регулярная составляющая в вероятностных моделях гораздо проще, чем в функциональных, а сами модели призваны описать 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.

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. Сейчас же стохастические модели широко используются для описания текстуры, реставрации изображений, для декомпозиции изображений на области и т.д.

Вероятностные методы также привлекаются для описания изображений не на пиксельном уровне, а на более высоких уровнях. Например, существуют вероятностные модели для контуров на изображениях, для структурных описаний и т.д. Но их, скорее, следует отнести к представлениям соответствующих уровней.

Помимо стохастических моделей, описывающих какой-то один из аспектов изображений, в статистическом анализе изображений ставится и

общая проблема, которая формулируется следующим образом: можно ли найти единую стохастическую модель для изображений?

Ответ на данный вопрос ищется в рамках порождающих стохастических моделей изображений. В таких моделях предполагается, что есть набор скрытых переменных  $\chi = (\chi_1, \dots, \chi_N)$  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  $f$  – это изображение.

Моделью конкретного изображения будет являться набор значений этих скрытых переменных, апостериорные вероятности которых можно оценить по правилу Байеса:

$$p(\chi | f) \sim p(f | \chi)p(\chi). \quad (1.7)$$

Приведем простейшую модель, в которой яркости отдельных пикселей распределены по одному и тому же нормальному закону, характеризующемуся двумя параметрами: (средним)  $a$  и (дисперсией)  $\sigma$ .

$$p(f | (a, \sigma)) = \prod_{x \in \Omega} p(f(x)) = \prod_{x \in \Omega} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{f(x)-a}{\sqrt{2\sigma}}\right)^2}. \quad (1.8)$$

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.

При построении таких моделей важным является учет информации о статистиках естественных сцен и об особенностях в строении систем зрения у животных и человека. Изучение этих вопросов составляет две отдельные, но тесно связанные ветви исследований в области статистического моделирования изображений, и им посвящена обширная литература. Экспериментально установленное существование четкой связи между организацией нейронов в естественных зрительных системах и статистиками естественных сцен показывает важность учета этой информации при построении систем компьютерного зрения.

Описанный выше подход неявно предполагает, что существует некоторое распределение безусловных вероятностей по изображениям:

$$p(f) = \int_{\chi} p(f | \chi) d\chi. \quad (1.9)$$

Очевидно,  $p(f)$  будет сильно отличаться для конкретных приложений (для очень ограниченных задач является возможным даже явно установить это распределение). Поэтому если и существует некая универсальная стохастическая модель изображений, то она должна опираться лишь на самые общие свойства физического мира, не обращаясь к информации о

конкретных объектах (и, естественно, ее использование в частных приложениях будет менее эффективным).

Скрытые переменные должны каким-то образом описывать содержание сцены. Если это содержание достаточно произвольно, то построить стохастическую модель изображений сцены, которая могла бы быть применима на практике, оказывается проблематичным. Это относится и к математическим представлениям вообще: в них приходится использовать определенные упрощения, так что они оказываются недостаточно выразительными для описания содержания сцен.

#### *Средний уровень: промежуточные символьные представления*

В качестве конечной цели интерпретации изображений часто рассматривается присвоение группам пикселей меток, соответствующих некоторым объектам реального мира, таким как дом, дерево, дорога и т.д. Однако если бы была прямая взаимосвязь между пиксельным и семантическим содержанием изображения, то эта задача была бы давно решена.

Как отмечалось, в качестве основных причин различия внешнего вида одного и того же объекта на разных изображениях можно назвать смену ракурса (пространственное преобразование), изменения освещения (преобразование интенсивности), смена типа сенсора, собственную изменчивость объекта (например, сезонные изменения на аэрокосмических фотографиях). Значит, для назначения меток необходимо построить описания изображений, инвариантные перечисленным преобразованиям. К сожалению, математические модели 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 сильно зависит от свойств объектов наблюдения.

Поскольку различные изображения соответствуют одному и тому же объекту, получению интерпретации изображения, инвариантной некоторым преобразованиям, также соответствует и уменьшение размерности данных. Одна из проблем заключается в том, чтобы построить инвариантные представления, потеряв при этом как можно меньше полезной информации.

Представления, решающие в той или иной степени эту задачу, принято выделять в отдельный класс представлений среднего уровня. Поскольку результатом их применения является неизобразительная информация, их также называют промежуточными символьными представлениями. Они обращаются к структуре изображения, то есть к взаимосвязям между различными пикселями, поэтому такие представления часто называют структурными.

Взаимосвязи между группами пикселей подразумевают привлечение статистик высоких порядков (их сложно оценить по изображениям) либо негладких функций (для них сложно определить аналитические модели), что

и вызывает сложности при использовании строгих математических моделей для описания структуры изображений.

В компьютерном зрении вместо использования абстрактных представлений, опирающихся на понятие инварианта, рассматривается проблема восстановления форм и местоположений объектов по изображениям, что также можно рассматривать в качестве инвариантных описаний. Однако процесс такого восстановления неоднозначен и должен опираться на некоторые промежуточные представления. С другой стороны, в этих промежуточных представлениях должны выделяться некоторые характерные элементы, соответствующие реальным физическим особенностям видимых поверхностей.

Д. Марром был сформулирован следующий список допущений об особенностях видимого мира, указывающих на необходимость введения ряда промежуточных представлений, таких как контуры, структурные элементы и их группы.

1. Поверхности как реальные объекты: весь видимый мир можно рассматривать как некоторую композицию гладких поверхностей, функции отражательной способности которых могут отличаться сложной пространственной структурой.

2. Иерархическая организация: пространственная организация функции отражательной способности некоторой поверхности часто порождается совместным воздействием целого ряда различных процессов, каждый из которых относится к отдельному уровню.

3. Подобие: объекты, появляющиеся на некоторой поверхности в результате некоторого процесса порождения отражательной способности, действующего на некотором определенном масштабном уровне, обычно обладают большим сходством по размерам, локальному контрасту, цвету и пространственной организации между собой, чем с другими объектами этой же поверхности.

4. Пространственная непрерывность: характерные объекты, возникающие на некоторой поверхности в результате действия какого-то одного процесса, помимо того, что они обладают «внутренним» подобием, часто образуют определенную пространственную организацию, принимающую вид кривых, прямых и, возможно, более сложных конфигураций.

5. Непрерывность нарушений непрерывности: геометрическое место разрывов по глубине или ориентации поверхности почти везде гладко.

6. Непрерывность движения: при наличии любого нарушения непрерывности движения более чем в одной точке, например, вдоль некоторой прямой, следует считать, что имеет место граница объекта.

Необходимость нескольких представлений промежуточного уровня также часто иллюстрируют, 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 знаки промежуточного представления, которые уже, в свою очередь, образуют

конкретные объекты. Подобная аналогия, однако, не говорит, сколько и каких должно быть промежуточных уровней. Мы рассмотрим три таких уровня: контурные представления, представления в виде непрямых структурных элементов и представления в виде составных структурных элементов, которые строятся последовательно. Кроме того, выделяется признаковый подход, в котором переход к описанию изображения осуществляется за один этап. Хотя признаки методы не вполне согласуются с общим устройством зрительной системы, они могут соответствовать каким-то отдельным ее блокам и, кроме того, являются полезными для решения частных практических задач.

### *Признаковые представления изображений*

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

Можно классифицировать различные признаки изображений следующим образом:

- 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.
- 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 вычисляются по некоторым промежуточным описаниям изображений.

Точки, в которых значение некоторого локального признака достигает экстремального значения, называются *точками интереса*. 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 с методами на основе структурных представлений, построение которых начинается с извлечения контуров.

### *Контурные представления изображений*

Под контуром обычно понимается местоположение локального изменения или резкого перепада яркости на изображении. Контурные служат простейшими признаками изображений, используемыми в целях анализа изображений для выделения границ объектов, и являются основой для построения структурных элементов. Хотя результат выделения контуров еще сложно назвать структурным описанием изображения, его также нельзя отнести и к низкоуровневым представлениям.

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 как границы этих областей либо как их срединные оси. Хотя такие представления весьма различны, они содержат идентичную информацию. Тем не менее, какие-либо из них могут быть менее удобными. Так, например, срединная ось обладает повышенной чувствительностью к малым возмущениям формы соответствующей области.

Локальные процедуры основываются либо на определении цепочек максимумов на градиентном поле (или пересечений нуля второй производной), либо на непосредственной аппроксимации яркостных переходов. Поскольку выделение локальных яркостных переходов может рассматриваться как вариант сегментации, то локальные и глобальные процедуры имеют много общего. Тем не менее, локальные процедуры существенно проще в реализации и более правдоподобны с нейрофизиологической точки зрения, поэтому они более популярны.

Построение градиентного поля полутонового изображения осуществляется с помощью операторов, выполняющих дискретное дифференцирование. Классическими операторами, служащими для этой цели, являются операторы Робертса, Превитта и Собела (а также другие операторы), которым соответствуют маски

$$H_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, H_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

$$H_1 = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, H_2 = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

$$H_1 = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, H_2 = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}.$$

На рис. 1.9 представлен пример построения градиентного поля изображения. Темным цветом показаны наибольшие значения. Как видно из рисунка, величина градиента максимальна на границах объектов.

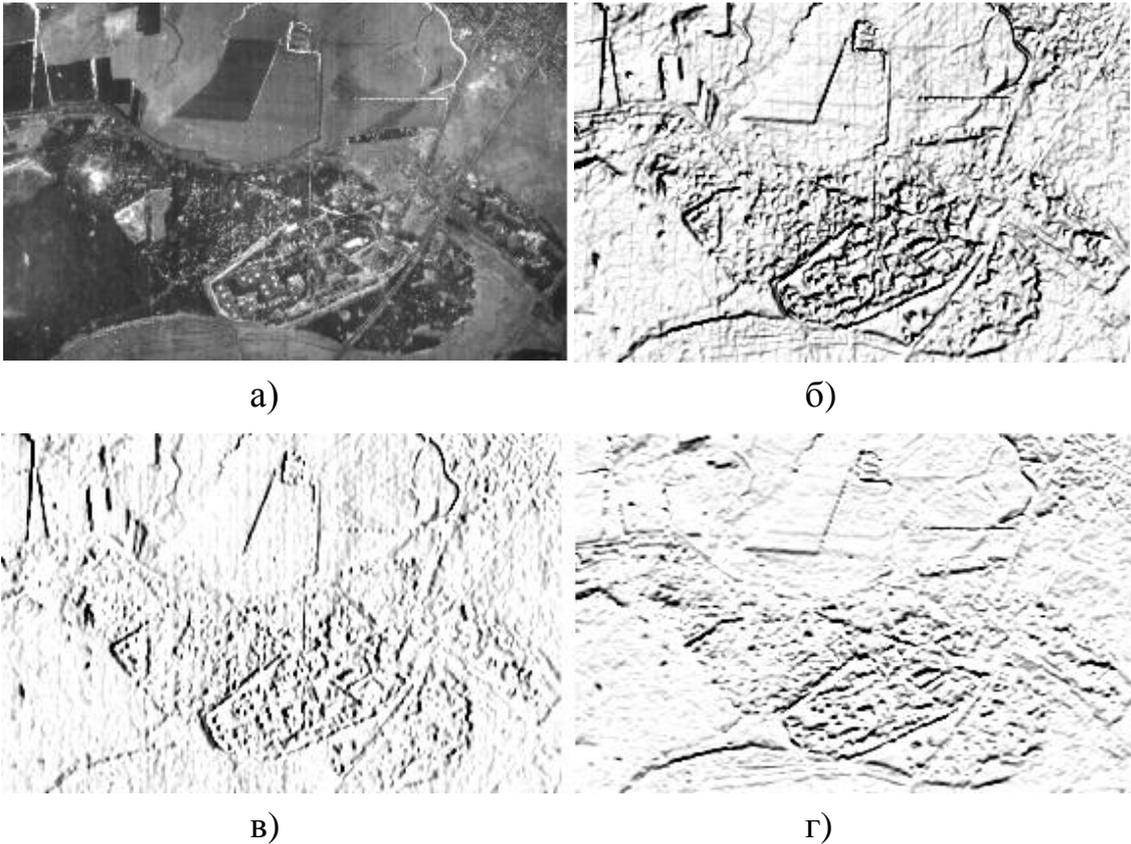


Рис. 1.9. Пример изображения (а), для которого выполнено численное дифференцирование с помощью оператора Собела: (б) – модуль градиента, (в) и (г) – горизонтальная и вертикальная компоненты градиента

Помимо указанных ранее масок 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. Помимо получения более гладких контуров, более удобных для дальнейшего анализа, привлечение масок разных размеров обосновывается иерархичностью организации физического мира.

Были предложены различные методы для формирования контуров на разных уровнях разрешения. В частности, Д. Марром было обосновано использование фильтра  $\nabla^2 G$ , где  $\nabla^2$  – оператор Лапласа и  $G = \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$  – Гауссова функция. Более того, выглядит правдоподобным, что именно эту функцию воспроизводят ганглиозные клетки сетчатки и клетки наружного коленчатого тела.

Сейчас популярны такие подходы обнаружения контуров, как, например, предложенные Канни, Дерешем или Линденбергом. В частности, на рис. 1.10 представлен пример выделения контуров на изображении 1.9а с помощью метода Дереша. Предлагается также и много других методов.

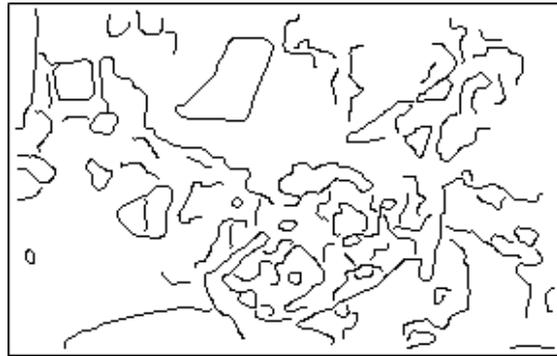


Рис. 1.10. Контур, выделенные с помощью фильтрации Дериша на изображении 1.9а

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 такой формализации понятия контура (о существовании этой проблемы свидетельствует разнообразие подходов к обнаружению контуров).

Как правило, приводятся следующие аргументы в пользу привлечения контуров:

- 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, условиях погоды 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 (более характерных для области компьютерного зрения, а не анализа изображений) аргументация использования контуров заключается в том, что нарушения непрерывности в физическом пространстве (поверхности) порождают также и перепады яркости на изображении. А значит, последние могут быть использованы для получения информации о реальных поверхностях.

Хотя все эти аргументы являются убедительными, они все же относятся к тем контурам, которые хотелось бы получить в результате работы детектора контуров, а не к привлекаемым в них моделям изображений. Таким образом, процедура перехода от низкоуровневых представлений изображений к контурным представлениям требует дальнейшей конкретизации.

Сами контурные представления также допускают использование строгой математической теории (например, контур может быть представлен в виде комплекснозначного сигнала). В рамках таких теорий контурное описание может рассматриваться как конечный результат процедуры интерпретации изображения и использоваться для дальнейшего анализа.

Популярным является использование контуров в целях совмещения пары изображений или изображения с векторной моделью (например, картой местности или чертежом детали) на основе преобразования расстояний, описание формы объектов или областей по их контурам, например, с помощью методов математической морфологии, для решения задачи стереозрения и т.д. Контурные представления также служат основой для построения структурных описаний изображений.

#### *Непроизводные структурные элементы*

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

В качестве основы структурных представлений изображений могут быть использованы точки интереса, строящиеся в рамках признакового подхода. 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”. Используются также детекторы признаков, предназначенные для узких предметных областей, например, для обнаружения лиц. Исследуется и противоположная возможность, заключающаяся в автоматическом построении детекторов признаков на основе содержащейся в изображении информации.

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 (см. рис. 1.11). 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.

Существует также и промежуточный подход, при котором выделяются геометрические элементы на основе краевых точек, не объединенных в связанные контуры. 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. 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 (Hough).

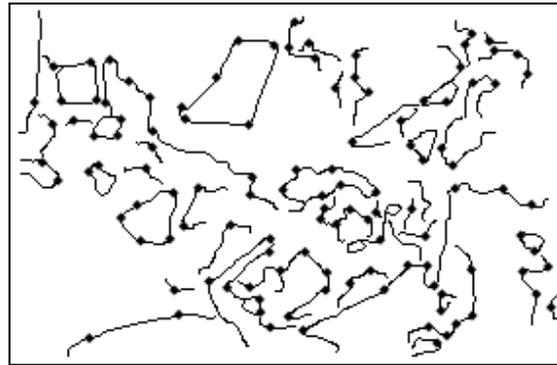


Рис. 1.11. Точки максимальной кривизны, выделенные на контурах, представленных на рис. 1.10

Разнообразие промежуточных символьных представлений еще больше, чем количество разных детекторов контуров. Строгие методы работы с такими представлениями существуют (например, на основе теории формальных грамматик или на основе теории графов), но в них предполагается, что исходное представление дано априори. Обоснования же конкретных методов построения представлений гораздо слабее.

#### *Составные структурные элементы*

Дальнейшее развитие структурного подхода заключается в формировании на основе непроектируемых 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. Последнее важно, например, в гео-информационных системах для извлечения изображения по содержанию.

Существуют различные подходы к группированию непроектируемых структурных элементов с целью образования более сложных структурных элементов, вид которых не зависит от предметной области. К таким элементам относятся, например, "L", "U" и другие соединения, пары параллельных прямых линий, пятна, составленные из пятен меньших размеров, и т.д. Однако все подходы к группированию в том или ином виде опираются на подобие структурных элементов или регулярность их расположения (к примеру, на рис. 1.12 группирование осуществляется именно на основе положения элементов). На основе подобия составные структурные элементы могут дальше группироваться, формируя более крупномасштабные элементы изображения.



Рис. 1.12. Иллюстрация к возможности группирования зрительной системой человека разнотипных элементов на основе регулярностей в их взаимном расположении, благодаря чему достигается возможность распознавания изображенного на рисунке объекта

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

*Верхний уровень: методы, основанные на знаниях*

Как уже отмечалось выше, в качестве конечной цели зрительного анализа часто рассматривают назначение семантических меток областям на изображении, то есть описание изображений сцен на естественном языке. Тогда в качестве конечного представления изображений рассматривается некоторая система представления знаний.

Более того, многие авторы указывают на необходимость использования знаний и в самом процессе анализа изображений. 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 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.

Простым примером неявного использования знаний может служить распознавание конкретных объектов на изображениях. И действительно, в этих задачах ведется целенаправленный поиск объектов, принадлежащих определенным (обычно очень узким) классам. Знание свойств этих объектов используется для их распознавания.

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,

коэффициентов Фурье или вейвлет преобразования и т.д. Распознаваемые объекты могут храниться в виде своих контурных или структурных 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: 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 интерпретации изображений: 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 экспертных систем и изучаются в области искусственного интеллекта.

Как правило, системы, основанные на знаниях, априорно имеют высокоуровневое описание сцены. 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 интерпретаций, каждая из которых выдвигается в качестве гипотезы и проверяется на соответствие с изображением. Это хотя и позволяет избежать комбинаторного взрыва числа возможных интерпретаций, характерного для подхода снизу вверх, но делает методы, основанные на знаниях, способными работать лишь в сильно ограниченных предметных областях.

Более того, при разработке систем, основанных на знаниях, также возникает потребность решать проблему построения промежуточных описаний изображений, то есть решать такие задачи, как сегментация, выделение контуров и структурных элементов и т.д. Таким образом, прежде чем рассматривать возможность создания универсальной системы зрения, основанной на знаниях, необходимо решить эту же проблему для построения промежуточного символьного описания.

Также понятно, что построение этих описаний в зрительной системе животных и человека первично по сравнению с построением высокоуровневых описаний сцен. Трудно утверждать, что младенцы обладают какими-либо семантическими знаниями об анализируемых их зрительной системой сценах. Хотя у человека действительно могут использоваться высокоуровневые знания для управления восприятием (например, этим объясняет т.н. эффект перцептивной готовности), но накопление этих знаний осуществляется на основе более низкоуровневых представлений изображений. К сожалению, в большинстве существующих систем компьютерного зрения знания о предметной области закладываются вручную. Это делает процесс создания таких систем трудоемким, а сами системы – узкоспециализированными.

Существуют исследовательские проекты, в которых система обучается построению семантического описания сцен за счет одновременного использования зрительного и лингвистического канала информации. Этот подход будет описан при обсуждении проблем понимания изображений.

### *Иерархические представления изображений*

Принято считать, что статистические решения о принадлежности некоторого сигнала с шумами к определенному классу должны основываться на как можно большем числе наблюдаемых отсчетов сигнала, и что промежуточные дискретные решения относительно подмножеств отсчетов вредны, так как они разрушают информацию; однако, как уже упоминалось, опыт решения задач анализа изображений доказывает целесообразность принятия подобных промежуточных решений.

Действительно, изображения могут трактоваться как зашумленный сигнал, отсчетами которого являются интенсивности отдельных пикселей (или клеток растра), а интерпретация изображений – как статистический вывод. Вывод, дающий наиболее надежные результаты, должен был бы опираться на исходные значения интенсивностей всех пикселей изображения. Почему же подобные методы анализа изображений (например, методы сравнения с эталоном в распознавании и корреляционные методы в совмещении изображений) оказались недостаточно эффективными?

Один из возможных ответов на этот вопрос можно сформулировать следующим образом. Проблема интерпретации является NP-полной, поэтому сложность нахождения точного решения этой проблемы возрастает не менее чем экспоненциально с увеличением размера изображения. Это ограничение является принципиальным, и оно подразумевает, что требуется искать субоптимальные методы интерпретации. Принятие промежуточных решений (или построение промежуточных представлений) и является искомым подходом к решению проблемы комбинаторного взрыва.

При исследовании существующих методов интерпретации изображений можно выделить два различных способа введения иерархичности в процесс анализа изображений: иерархичность по пространственному масштабу (*методы с переменной разрешающей способностью*) и иерархичность по

уровням абстрактности привлекаемых представлений (многоуровневые методы).

Существуют два основных аргумента в пользу привлечения методов с переменной разрешающей способностью. Первый аргумент апеллирует к природе изображений: изображения не содержат предустановленного характерного масштаба в отличие, например, от речевой или тактильной информации. В более широкой формулировке иерархичность пространственной организации постулируется как одно из основных общих свойств видимого мира, которое должно учитываться при разработке представлений изображений. В качестве более практического аргумента приводится возможность с помощью методов с переменной разрешающей способностью заметно увеличивать производительность алгоритмов: приближенное решение сначала находится на основе изображений с уменьшенным разрешением, а затем это решение лишь уточняется при постепенном увеличении разрешения, вместо того, чтобы решать полную задачу на наиболее детальном уровне.

Очевидно, вероятность получить неоптимальное решение возрастает, поскольку при нахождении начального приближения используется не вся имеющаяся информация, однако при этом достигается значительный выигрыш в производительности, поэтому при одинаковом времени работы иерархических и неиерархических методов результат первых оказывается лучше.

Представления, использующие несколько масштабных уровней, существуют для различных уровней абстракции, хотя наиболее исследованными являются низкоуровневые представления, такие как пирамидально-рекурсивные структуры, вейвлет и фрактальные представления и т.д. Целесообразность построения контурных и структурных представлений на нескольких масштабных уровнях была указана сравнительно давно, и в некоторых подходах такие представления в определенной степени используются, однако сами представления детально не исследовались. При построении представлений, основанных на знаниях, пространственная иерархичность может осуществляться с помощью отношений «часть-целое».

Другой тип иерархичности – по уровням абстракции – имеет еще более слабое обоснование. Аргументы, выдвигаемые в защиту необходимости вводить промежуточные представления, упоминались ранее. Как правило, они носят чисто эмпирический характер. Многоуровневыми методами обычно называют методы, привлекающие одновременно несколько представлений. В таких методах сначала последовательно строятся представления возрастающих уровней абстракции, затем задача анализа изображений решается на наиболее высоком уровне с последующим уточнением на более низких уровнях.

Здесь так же, как и в методах с переменной разрешающей способностью, сначала находится грубое решение на основе данных с уменьшенной размерностью, а затем оно уточняется. Есть, однако, и существенное

различие: понижение размерности данных здесь осуществляется при выделении наиболее релевантной информации или инвариантных признаков, в то время как в методах переменной разрешающей способности понижение размерности данных осуществляется независимо от содержания изображения.

Как отмечалось ранее, принятие промежуточных решений приводит к потере информации. В связи с этим при построении многоуровневых представлений изображений важной проблемой оказывается введение обратных связей от верхних уровней к нижним, которые бы позволяли восстанавливать потерянную информацию и устранять неоднозначность интерпретации. Теория введения подобных связей в области компьютерного зрения не достаточно разработана.

Некоторые данные были получены при исследовании зрительной системы человека. Общая идея здесь заключается в том, что, получив интерпретацию исходных данных на более высоком уровне, мы можем уточнить более низкоуровневую интерпретацию. К примеру, при чтении рукописного текста, написанного не слишком хорошим подчерком, в некотором слове человек для одной из букв может иметь несколько гипотез ее значения. Правильная гипотеза позволит получить осмысленное слово (элемент представления следующего уровня) и будет поддержана за счет положительной обратной связи с этого уровня. Другие же гипотезы, не дающие хороших интерпретаций на следующем уровне, будут подавлены отрицательной обратной связью. Правильные интерпретации на различных уровнях как бы входят в резонанс, что позволяет бороться с неоднозначностями, возникающими в подходе снизу вверх.

### **Вопросы и упражнения**

1. Поясните понятие представления и его роль в компьютерном зрении.
2. Какие проблемы позволяют решить низкоуровневые представления изображений?
3. Какие типы представлений можно отнести к промежуточным символьным представлениям?
4. В чем сущность подхода на основе признаков?
5. Каковы основные аргументы в пользу привлечения контуров в анализе изображений?
6. Какие типы структурных представлений вы знаете?
7. Что такое производные и составные структурные элементы?
8. Для каких представлений характерен подход «сверху вниз», а для каких – «снизу вверх»?
9. Чем отличаются методы с переменным разрешением и многоуровневые методы?

## 2. Сопоставление изображений

### 2.1. Геометрические преобразования изображений

Одной из базовых задач анализа изображений является их *сопоставление*, то есть отождествление *сопряженных точек*. Сопряженными точками на двух изображениях называются точки, являющиеся проекциями одной и той же точки сцены. Сопоставление изображений может быть необходимо как для распознавания изображений, так и для их *совмещения* (то есть такого взаимного преобразования изображений, при котором координаты сопряженных точек будут совпадать).

Задача сопоставления возникает в связи с тем, что изображения получаются с разных ракурсов, и одна и та же точка сцены может обладать разными координатами на изображениях. Изменение ракурса – одна из наиболее типичных причин изменчивости изображения. Следует сразу отметить, что изменение координат сопряженных точек на изображениях (что можно охарактеризовать как *геометрические искажения*) может быть вызвано не только изменением ракурса съемки, но и собственной изменчивостью объектов.

В компьютерном зрении сопоставление изображений оказывается необходимым как для их распознавания, так и для оценки взаимного расположения камер в момент съемки двух изображений (что может использоваться для определения ориентации камеры, определения параметров ее движения и оценивания расстояний до объектов). Чтобы определить, как смена ракурса влияет на расположение сопряженных точек на паре изображений, необходимо знать, как именно происходит формирование изображения оптической системой.

Прежде чем переходить к описанию методов сопоставления изображений, рассмотрим наиболее типичные случаи: центральную и параллельную проекции, что даст представление о возможных геометрических преобразованиях изображений.

#### *Центральная и ортогональная проекции*

Рассмотрим процесс формирования изображения на примере *центральной (перспективной) проекции*, создаваемой простейшей оптической системой – камерой-обскура, то есть камерой с точечным отверстием, расположенным на некотором расстоянии от плоскости изображения (см. рис. 2.1). При этом лучи, отраженные от точек трехмерной сцены, проецируются на плоское изображение, проходя через отверстие. С данной камерой можно связать систему координат, начало которой совпадает с отверстием (рис. 2.2), а ось  $Z$  направлена вдоль оптической оси (перпендикулярно плоскости изображения).

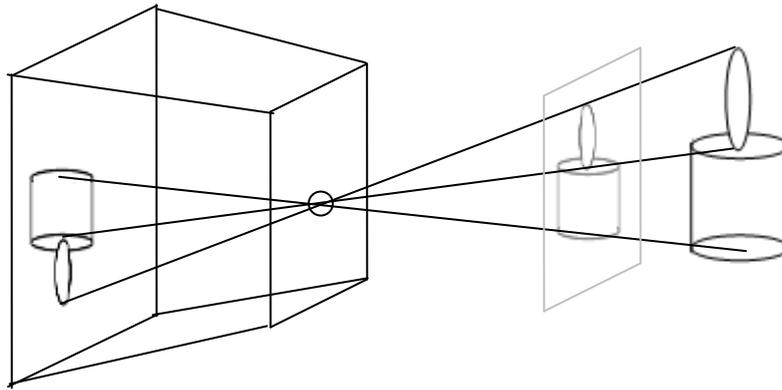


Рис. 2.1. Формирование перевернутого изображения (которому соответствует мнимое неперевернутое изображение) объекта в камере-обскура

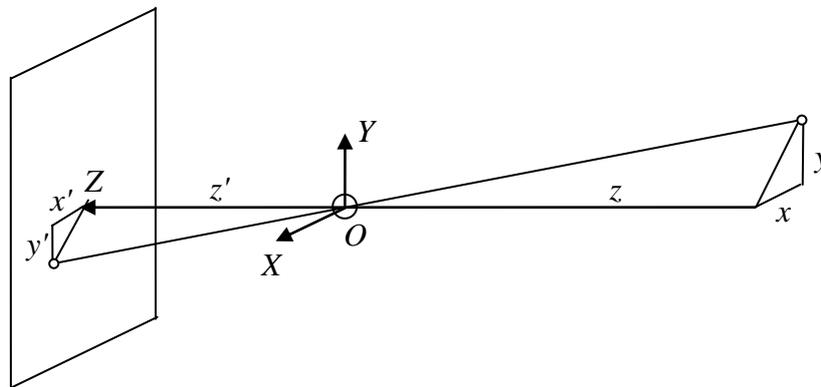


Рис. 2.2. Система координат при центральной проекции

Несложно убедиться, что выполняются следующие соотношения между координатами точек сцены и соответствующих им точек изображения:  $x'/z' = x/z$  и  $y'/z' = y/z$ . Поскольку координата  $z'$  постоянна для всех точек, обозначим ее через  $f'$  – расстояние от отверстия до плоскости изображения. Тогда соотношения координат примут вид

$$x'/f' = x/z \text{ и } y'/f' = y/z. \quad (2.1)$$

При этом идеальная линза дает тот же тип проекции, что и камера-обскура (см. рис. 2.3), в связи с чем центральная проекция является крайне распространенной в системах компьютерного зрения.

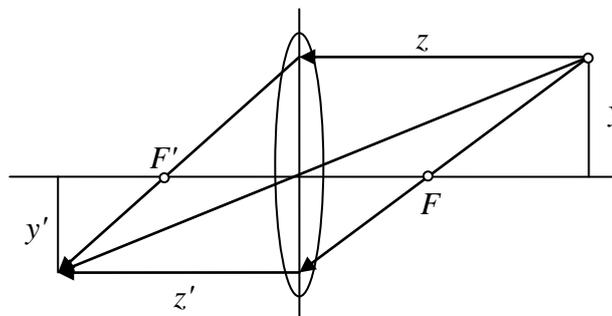


Рис. 2.3. Формирование изображения тонкой линзой

Естественно, реальные линзы обладают рядом aberrаций, с которыми на практике приходится бороться, однако компенсация этих aberrаций обычно

выполняется до решения задач компьютерного зрения, в связи с чем этот вопрос мы здесь рассматривать не будем. В целом, центральная проекция является вполне удовлетворительным приближением к процессу формирования изображений во многих задачах.

Как видно из уравнений (2.1), координаты проекции объекта (и, соответственно, размеры изображения объекта) зависят от удаленности до этого объекта. Важной характеристикой сцены является ее *глубина* – разброс расстояний от видимых точек до камеры. Если мы будем рассматривать некоторую сцену, расстояние до точек которой примерно постоянно (то есть расстояние до сцены значительно больше глубины сцены), то, положив  $m = -f'/z$ , получим

$$x' = -mx \text{ и } y' = -my, \quad (2.2)$$

где  $m$  – коэффициент увеличения (знак минус был взят, чтобы этот коэффициент был положительным, поскольку  $z < 0$ , а  $f' > 0$ , однако на практике удобнее рассматривать координаты проекции объектов на мнимом изображении, и тогда знак минус пропадет). Проекция, описываемая уравнениями (2.2), называется *слабоперспективной*, поскольку в этом случае отсутствует эффект перспективы (уменьшение видимых размеров с увеличением расстояния).

Для удобства координаты в плоскости мнимого изображения можно нормировать таким образом, чтобы коэффициент умножения был равен 1. В этом случае получим уравнение *ортогональной* (или *параллельной*) проекции (см. рис. 2.4)

$$x' = x \text{ и } y' = y. \quad (2.3)$$

Таким образом, при ортогональной проекции просто теряется информация о координате  $z$ . Данная проекция называется ортогональной, поскольку обычно возникает при съемке необъемной сцены, плоскость которой перпендикулярна оптической оси.

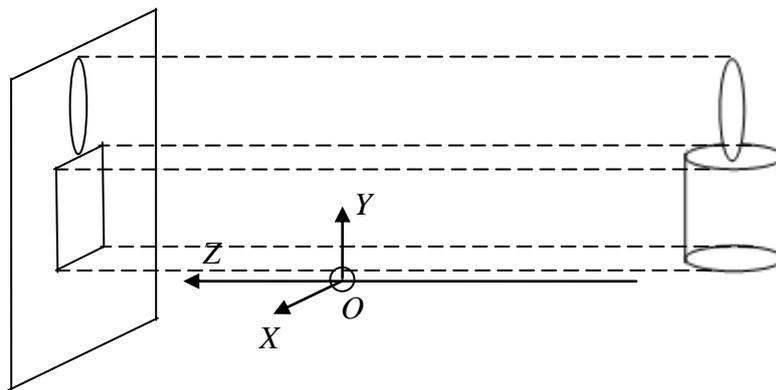


Рис. 2.4. Формирование изображения в случае ортогональной проекции

#### *Взаимное пространственное преобразование изображений*

Посмотрим, как меняются координаты проекции точки на изображении при изменении ракурса съемки. При этом может меняться положение начала координат  $O$ , а также направление осей (см. рис. 2.5).

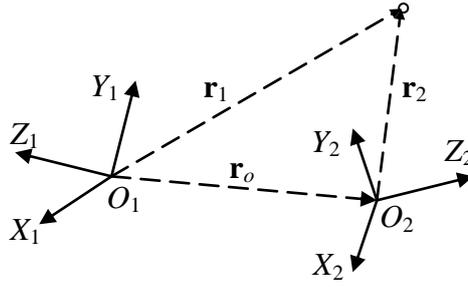


Рис. 2.5. Смена ракурса съемки

Переход от одной системы координат к другой описывается *уравнением абсолютной ориентации*

$$\mathbf{r}_2 = \mathbf{R}\mathbf{r}_1 + \mathbf{r}_o, \quad (2.4)$$

где  $\mathbf{R} = (r_{i,j})_{i,j=1}^3$  – матрица поворота,  $\mathbf{r}_o = (\Delta x, \Delta y, \Delta z)$  – вектор переноса. Матрица поворота получается путем перемножения матриц вращения вокруг каждой оси. В частности, вращение вокруг оси  $Z$  на угол  $\varphi$  описывается матрицей

$$\mathbf{R}_Z = \begin{pmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Обратим внимание, что помимо ракурса съемки может также меняться и параметр  $f'$ .

В общем случае для получения связи между координатами сопряженных точек двух изображений необходимо воспользоваться уравнением проекции (2.1) и преобразованием координат (2.4), при этом исключив физические координаты точек.

Перепишем уравнение центральной проекции (2.1) для первой камеры в виде

$$\begin{aligned} x_1 &= z_1 x'_1 / f'_1, \\ y_1 &= z_1 y'_1 / f'_1. \end{aligned}$$

Подставим эти значения в уравнения перехода между системами координат двух камер (2.4)

$$\begin{aligned} x_2 &= r_{11}x_1 + r_{12}y_1 + r_{13}z_1 + \Delta x = r_{11}z_1 x'_1 / f'_1 + r_{12}z_1 y'_1 / f'_1 + r_{13}z_1 + \Delta x, \\ y_2 &= r_{21}x_1 + r_{22}y_1 + r_{23}z_1 + \Delta y = r_{21}z_1 x'_1 / f'_1 + r_{22}z_1 y'_1 / f'_1 + r_{23}z_1 + \Delta y. \end{aligned}$$

Теперь запишем уравнение центральной проекции для второй камеры:

$$\begin{aligned} x_2 &= z_2 x'_2 / f'_2, \\ y_2 &= z_2 y'_2 / f'_2. \end{aligned}$$

Подставив выражения для  $x_2$  и  $y_2$ , получим

$$\begin{cases} (r_{11} x'_1 / f'_1 + r_{12} y'_1 / f'_1 + r_{13}) z_1 + \Delta x = z_2 x'_2 / f'_2, \\ (r_{21} x'_1 / f'_1 + r_{22} y'_1 / f'_1 + r_{23}) z_1 + \Delta y = z_2 y'_2 / f'_2. \end{cases} \quad (2.5)$$

Координату  $z_2 = r_{31}x_1 + r_{32}y_1 + r_{33}z_1 + \Delta z$  также можно выразить через координату  $z_1$  (исключив  $x_1$  и  $y_1$  с помощью уравнения центральной проекции), однако, как видно, в этих уравнениях невозможно исключить дальность  $z_1$  до точки сцены. Таким образом, не существует однозначного отображения между сопряженными точками двух изображений, снятых с разных ракурсов, если дальности до этих точек неизвестны.

Можно, однако, рассмотреть достаточно часто встречающийся частный случай, при котором сцена является плоской, то есть имеется связь координат:  $z_1 = ax_1 + by_1 + c$ , где  $a$ ,  $b$  и  $c$  – неизвестные, но постоянные коэффициенты. Также для простоты положим, что  $z_2 = \mathbf{const}$ , то есть посмотрим, как преобразуется изображение плоской сцены, снятой под наклоном и ортогонально. Из условия  $z_1 = ax_1 + by_1 + c$  и связи координат для первого уравнения в системе (2.5) имеем

$$z_1 = a(z_1 x'_1 / f'_1) + b(z_1 y'_1 / f'_1) + c \Rightarrow z_1 = c / (1 - a x'_1 / f'_1 - b y'_1 / f'_1).$$

Тогда

$$c(r_{11} x'_1 / f'_1 + r_{12} y'_1 / f'_1 + r_{13}) / (1 - a x'_1 / f'_1 - b y'_1 / f'_1) + \Delta x = z_2 x'_2 / f'_2,$$

откуда

$$x'_2 = \frac{(f'_2 c / z_2)(r_{11} x'_1 / f'_1 + r_{12} y'_1 / f'_1 + r_{13} + \Delta x(1 - a x'_1 / f'_1 - b y'_1 / f'_1))}{1 - a x'_1 / f'_1 - b y'_1 / f'_1}.$$

Поскольку нас не интересуют сами значения  $r_{i,j}$ ,  $z_2$ ,  $a$ ,  $b$  и  $c$  (которые при сопоставлении изображений являются неизвестными; кроме того, неизвестными могут быть также значения  $f'_1$ ,  $f'_1 f'_2$ ) мы можем ввести замены, обозначив

$$a_0 = (f'_2 c / z_2)(r_{11} / f'_1 - a \Delta x / f'_1);$$

$$a_1 = (f'_2 c / z_2)(r_{12} / f'_1 - b \Delta x / f'_1);$$

$$a_2 = (f'_2 c / z_2)(r_{13} + \Delta x);$$

$$a_3 = -a / f'_1; \quad a_4 = -b / f'_1,$$

что даст выражение

$$x'_2 = \frac{a_0 x'_1 + a_1 y'_1 + a_2}{a_3 x'_1 + a_4 y'_1 + 1}. \quad (2.6a)$$

Аналогично можно получить

$$y'_2 = \frac{a_5 x'_1 + a_6 y'_1 + a_7}{a_3 x'_1 + a_4 y'_1 + 1}. \quad (2.6б)$$

Параметры  $a_i$  неизвестны и зависят от многих факторов. Однако мы видим, что в случае плоской сцены, даже снятой под некоторым углом и с близкого расстояния, координаты сопряженных точек связаны единообразно. Данные уравнения задают *проективное* преобразование плоскости.

Стоит отметить, что проективным преобразованием описывается не только переход от изображения сцены, снятой под углом, к изображению ортогонально снятой сцены, но и переход между двумя изображениями плоской сцены, снятой под произвольными ракурсами. Это можно

обосновать, если показать, что суперпозиция двух проективных преобразований также дает проективное преобразование, и обратное к проективному преобразование также является проективным.

Итак, изображения плоской сцены, снятой под произвольными ракурсами и с произвольных расстояний, в случае центральной проекции связаны единым проективным преобразованием.

Несложно показать, что в случае слабоперспективной проекции преобразование будет иметь более простой вид. Действительно, если в уравнениях (2.5) положить  $z_1 = \mathbf{const}$ , то после переобозначения неизвестных постоянных величин несложно получить уравнения *аффинного* преобразования

$$\begin{cases} x'_2 = a_0 x'_1 + a_1 y'_1 + a_2, \\ y'_2 = a_3 x'_1 + a_4 y'_1 + a_5. \end{cases} \quad (2.7)$$

Аналогичный результат можно было бы получить, подставив соотношения  $x'_1 = -m_1 x_1$  и  $y'_1 = -m_1 y_1$  в уравнения связи систем координат камер.

В случае ортогональной проекции, при которой плоскость сцены перпендикулярна оптической оси, вокруг которой может вращаться камера, изображения будут связаны преобразованием из группы *подобия*, которое является частным случаем аффинного преобразования при,  $a_0 = a_3 = a \cos \varphi$ ,  $a_1 = a \sin \varphi$ ,  $a_2 = -a \sin \varphi$ , где  $a$  – взаимный масштаб, а  $\varphi$  – угол поворота двух изображений. Если же сцена снимается с одного и того же расстояния при неизменном расстоянии от начала координат до плоскости изображения, то коэффициент взаимного масштаба будет равен 1, а преобразование изображений будет относиться к группе *движения*. В самом простом случае изображения будут отличаться только смещением  $x'_2 = x'_1 + \Delta x$ ,  $y'_2 = y'_1 + \Delta y$ .

### *Свойства пространственных преобразований изображений*

Важной характеристикой преобразований является то, как они искажают простейшие геометрические элементы. Посмотрим, к примеру, что происходит с прямой линией при проективном преобразовании (далее штрихи у координат в плоскости изображения будем опускать). Пусть линия на втором изображении задана уравнением  $ax_2 + by_2 + c = 0$ . Подставляя сюда уравнения проективного преобразования (2.6), получим

$$a \frac{a_0 x_1 + a_1 y_1 + a_2}{a_3 x_1 + a_4 y_1 + 1} + b \frac{a_5 x_1 + a_6 y_1 + a_7}{a_3 x_1 + a_4 y_1 + 1} + c = 0.$$

Для всех точек, в которых знаменатель отличен от нуля, получаем:

$$a(a_0 x_1 + a_1 y_1 + a_2) + b(a_5 x_1 + a_6 y_1 + a_7) + c(a_3 x_1 + a_4 y_1 + 1) = 0,$$

откуда

$$(aa_0 + ba_5 + ca_3)x_1 + (aa_1 + ba_6 + ca_4)y_1 + (aa_2 + ba_7 + c) = 0, \quad (2.8)$$

что также является уравнением линии.

Если мы на втором изображении зададим линию, параллельную данной,  $ax_2 + by_2 + c' = 0$ , то, как не сложно убедиться, коэффициенты при  $x_1$  и  $y_1$  в уравнении (2.8) изменятся.

Таким образом, при проективном преобразовании прямая линия остается прямой, однако параллельность линий нарушается. Это говорит о том, что проективным преобразованием (а, значит, при центральной проекции) квадрат отображается в произвольный четырехугольник.

Очевидно, при преобразованиях более частных видов (аффинном, подобия, движения и сдвига) прямые линии также останутся прямыми. Однако этими преобразованиями параллельность линий сохраняется. Действительно, для аффинного преобразования образ линии  $ax_2 + by_2 + c = 0$  на первом изображении примет форму

$$(aa_0 + ba_3)x_1 + (aa_1 + ba_4)y_1 + (aa_2 + ba_5 + c) = 0.$$

Коэффициенты при  $x_1$  и  $y_1$  для параллельной ей линии  $ax_2 + by_2 + c' = 0$  не изменятся, а изменится лишь расстояние между линиями.

Таким образом, при аффинном преобразовании сохраняется и параллельность линий (но не сохраняются углы между непараллельными линиями, в чем также несложно убедиться). Иными словами, аффинным преобразованием квадрат отображается в параллелограмм.

Более простым является преобразование подобия, которое сохраняет углы между линиями и отображает квадрат в квадрат.

Покажем, что проективное преобразование переводит окружность, заданную уравнением  $x_2^2 + y_2^2 = R^2$ , в эллипс

$$\left( \frac{a_0x_1 + a_1y_1 + a_2}{a_3x_1 + a_4y_1 + 1} \right)^2 + \left( \frac{a_5x_1 + a_6y_1 + a_7}{a_3x_1 + a_4y_1 + 1} \right)^2 = R^2.$$

Несложно убедиться, что после домножения на знаменатель будет получено квадратичное уравнение.

Аналогичный результат можно получить и для преобразования эллипса, причем не только для проективного, но и для аффинного преобразования. Окружность переводится в окружность только преобразованием подобия или преобразованиями более частных групп.

Таким образом, изображение плоского круглого диска является эллипсом, если только съемка не производится ортогонально. Несколько сложнее обосновать тот факт, что изображение сферы при центральной проекции также является эллипсом, но только если сфера находится очень близко к камере (на расстоянии, сопоставимом с ее радиусом). Уже на расстоянии, равном нескольким диаметрам, изображение сферы мало отличается от круга (изображение сферы также является кругом, если она расположена на оптической оси).

### *Типы пространственных преобразований изображений*

Геометрическое преобразование изображений, вызванное сменой ракурса и задающее связь координат сопряженных точек, может иметь

разный вид. Если это преобразование имеет фиксированный набор параметров и может быть единообразно применено к каждой точке изображения (когда эти параметры известны), то говорится, что взаимное пространственное преобразование изображений является *глобальным*.

Глобальные преобразования характерны для изображений плоских сцен, таких как, например, аэрокосмические изображения или изображения ленты конвейера (см. рис. 2.6). Наиболее часто встречающимися классами глобальных преобразований являются

- группа смещения,
- группа движения,
- группа подобия,
- аффинные преобразования,
- проективные преобразования.

Наиболее общими являются проективные преобразования, но также в некоторых задачах изображения преобразованы билинейными и полиномиальными преобразованиями, не сводящимися к проективным. Аффинные преобразования возникают при малой глубине сцены по сравнению с расстоянием до нее. Преобразования более простых типов появляются тогда, когда есть какие-то дополнительные ограничения.

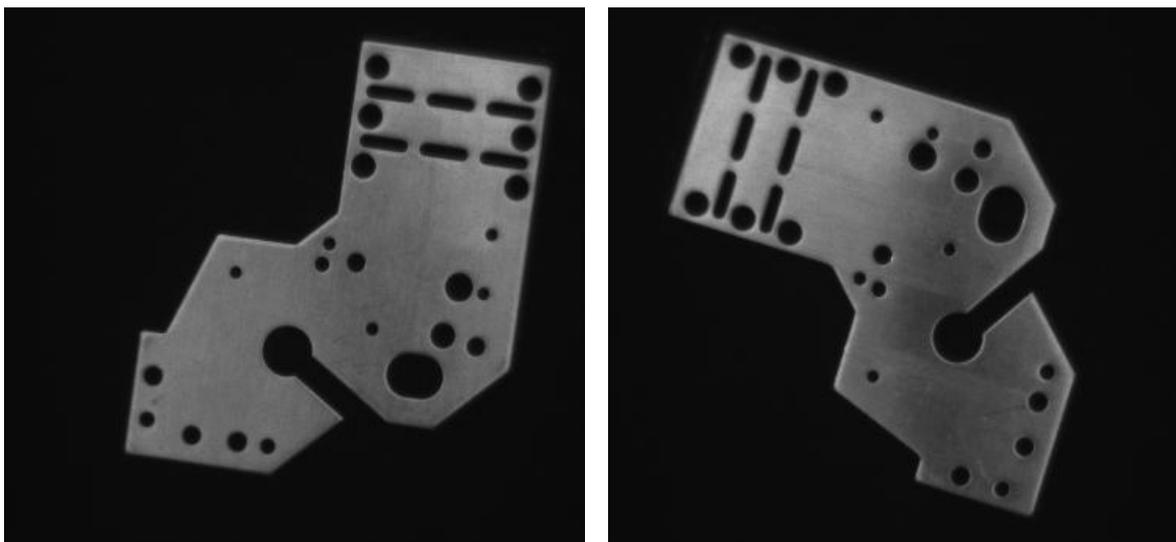


Рис. 2.6. Изображения детали на ленте конвейера, снятые с разных ракурсов; взаимное пространственное преобразование изображений является глобальным преобразованием движения

Часто, однако, сцены не являются абсолютно плоскими, в связи с чем взаимное пространственное преобразование изображений не описывается одной глобальной моделью, и оказывается необходимым привлекать модели *локальных преобразований*. Можно выделить следующие способы задания локальных пространственных преобразований

- кусочно-глобальные,
- глобальные с гладкой зависимостью параметров от координат на изображении,
- эластичные,
- поля смещений.

Кусочно-глобальные преобразования естественным образом возникают, если сцену представить состоящей из совокупности плоских поверхностей. В этом случае каждая область изображения, соответствующая некоторой поверхности, будет подвержена глобальному преобразованию (аффинному или проективному) с собственными значениями параметров. То есть параметры глобального преобразования будут меняться дискретным образом от области к области (рис. 2.7). К примеру, для случая кусочно-глобального аффинного преобразования можно записать

$$\begin{cases} x_2 = a_0(x_1, y_1)x_1 + a_1(x_1, y_1)y_1 + a_2(x_1, y_1), \\ y_2 = a_3(x_1, y_1)x_1 + a_4(x_1, y_1)y_1 + a_5(x_1, y_1). \end{cases} \quad (2.9)$$

При этом параметры преобразования как функции координат являются ступенчатыми.

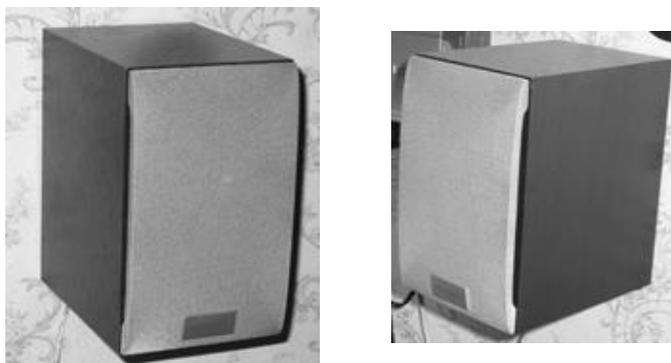


Рис. 2.7. Пример объекта, снятого с двух ракурсов; взаимное пространственное преобразование изображений может быть представлено как кусочно-глобальное

Если сцена состоит из гладких неплоских поверхностей, то поверхности в каждой точке такой сцены могут быть аппроксимированы плоскостью, однако наклон этой плоскости будет плавно меняться от точки к точке. В связи с этим и параметры глобального преобразования (задаваемого в той же форме (2.9)) будут также постоянно, но плавно меняться.

В некоторых приложениях (особенно, биомедицинских) сами объекты являются гибкими и отдельные их части могут смещаться, при этом оставаясь соединенными с другими частями объекта. Для описания пространственного преобразования между сопряженными точками изображений подобных объектов используются т.н. эластичные преобразования, которые, однако, являются достаточно специфичными, и мы на них здесь не будем останавливаться.

Наиболее общим (и потому недостаточно мощным в частных случаях) способом задания взаимного пространственного преобразования изображений являются поля смещений, в которых каждой точке изображения приписывается собственное смещение, переводящее ее в сопряженную точку на другом изображении. Поле смещений может пониматься и более узко, как локальное отклонение координат сопряженных точек от глобального преобразования. К примеру, поле локальных смещений может возникнуть на

аэрокосмическом изображений с достаточно высокими зданиями или холмистой местностью. Как видно из рис. 2.8, у крыши каждого здания имеется собственное (относительно небольшое) смещение, которое не компенсируется единым глобальным преобразованием, относящимся в данном случае к аффинной группе.



Рис. 2.8. Пример аэрокосмических снимков, взаимное геометрическое преобразование которых может быть описано как аффинное для объектов, находящихся на земле, с дополнительным полем локальных смещений для крыш зданий

Таким образом, изменение ракурса съемки приводит к геометрическим искажениям изображений, которые могут быть разных типов. При отождествлении сопряженных точек на двух изображениях необходимо учитывать то, как именно могут быть преобразованы изображения. Помимо этого, проблема определения параметров преобразования почти полностью эквивалентна проблеме сопоставления изображений. Естественно, чем сложнее допустимые пространственные преобразования, тем более сложной

является проблема сопоставления. Эта проблема может решаться с привлечением различных представлений изображений. При этом с каждым представлением связаны собственные методы сопоставления (см. рис. 2.9).

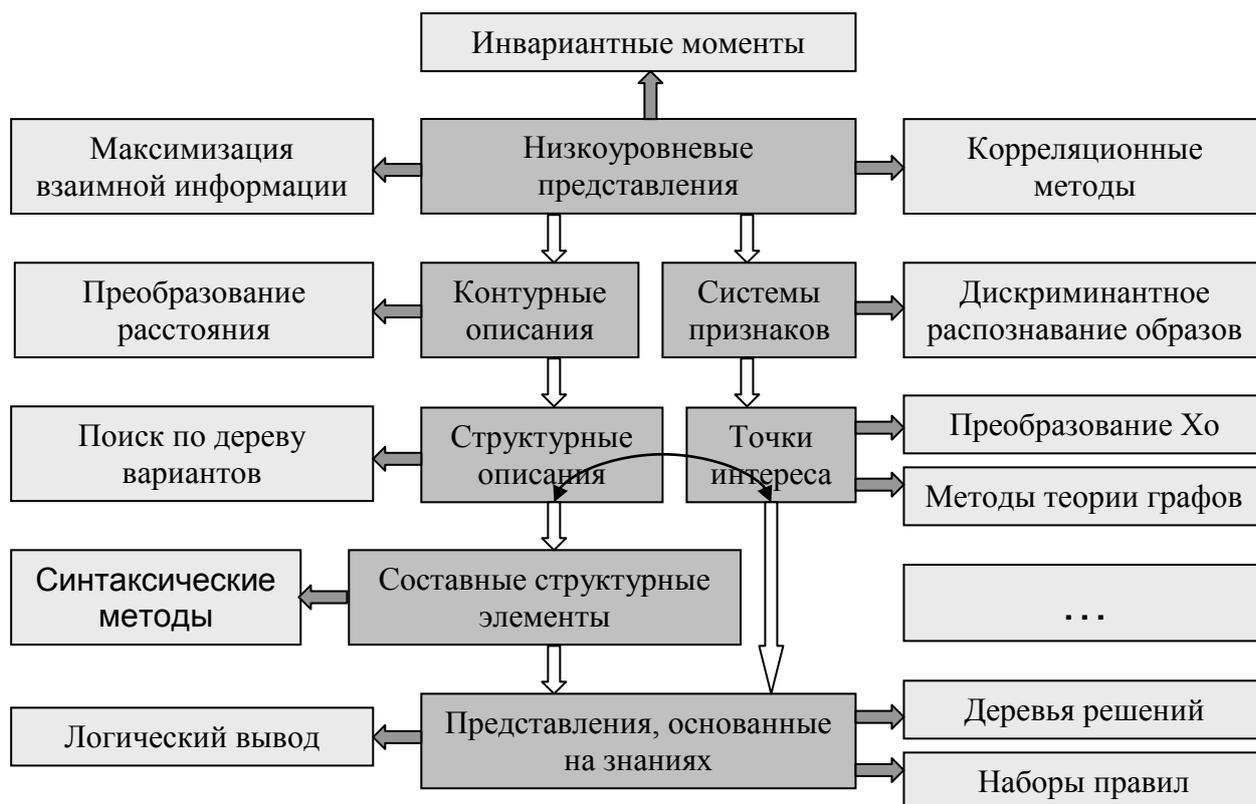


Рис. 2.9. Разнообразие методов сопоставления изображений в зависимости от привлекаемых представлений

Совместно с низкоуровневыми представлениями используются и такого же рода меры сходства изображений, основанные на понятиях корреляции, взаимной информации, статистических моментах и т.д. Если изображение описывается вектором признаков, то для его распознавания может использоваться богатый аппарат дискриминантного распознавания образов. В контурных методах мерой сходства служит, как правило, среднее расстояние между контурами двух изображений, для вычисления которого выполняется преобразование расстояния. При сопоставлении описаний изображений, включающих структурные элементы или точки интереса, могут использоваться разнообразные методы от поиска по дереву вариантов возможных соответствий и синтаксического вывода до весьма специфичных, таких как преобразование Хафа. В методах сопоставления, основанных на знаниях, используется аппарат манипулирования знаниями, соответствующий выбранному представлению. Рассмотрим некоторые из методов сопоставления, начиная с низкоуровневых.

### Вопросы и упражнения

1. Что такое сопряженные точки изображений?
2. В чем заключаются задачи сопоставления и совмещения изображений?

3. Докажите, что суперпозиция двух проективных преобразований, а также преобразование, обратное к проективному, являются также проективными преобразованиями.
4. Докажите, что при аффинном преобразовании углы между непараллельными линиями могут изменяться.
5. Докажите, что преобразование подобия сохраняет значения углов.
6. Докажите, что аффинное преобразование в общем случае переводит окружность в эллипс.
7. Докажите, что преобразование подобия переводит окружность в окружность.
8. Какие типы пространственных преобразований изображений вы знаете?
9. При каких условиях изображения одной и той же сцены, снятой с разных ракурсов связаны единым глобальным преобразованием?
10. Какое пространственное преобразование изображений плоской сцены возникает при центральной проекции?

## 2.2. Корреляционные методы сопоставления изображений

### *Амплитудно-фазовая корреляция*

Проблема сопоставления изображений может решаться с привлечением разных типов представлений, начиная от математических и заканчивая представлениями знаний. Классическими методами сопоставления являются корреляционные методы, базовая идея которых крайне проста и заключается в попиксельном сравнении изображений, хотя различия в реализациях могут быть достаточно значительны.

В простейшем случае мера сходства двух изображений как функция взаимного смещения  $(\Delta x, \Delta y)$  вычисляется как квадрат среднеквадратичного отклонения яркостей соответствующих (с учетом величины смещения) пикселей

$$\Sigma_{f_1, f_2}(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (f_1(x, y) - f_2(x - \Delta x, y - \Delta y))^2, \quad (2.10)$$

где  $f_1$  и  $f_2$  – изображения одного и того же объекта, отличающиеся только сдвигом;  $N \times N$  – размер области, в которой вычисляется корреляция.

Как видно из формулы (2.10), данный подход позволяет определять взаимные смещения двух изображений при отсутствии других геометрических преобразований, а также яркостных изменений.

Величину  $\Sigma_{f_1, f_2}$  несложно преобразовать к виду

$$\begin{aligned} \Sigma_{f_1, f_2}(\Delta x, \Delta y) &= \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1^2(x, y) - \frac{2}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1(x, y) f_2(x - \Delta x, y - \Delta y) + \\ &+ \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_2^2(x - \Delta x, y - \Delta y). \end{aligned}$$

Полагая величины  $\frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1^2(x, y)$  и  $\frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_2^2(x - \Delta x, y - \Delta y)$  не зависящими от смещения (что верно при некоторых оговорках), можно получить, что минимизация критерия СКО (2.10) эквивалентна максимизации значения корреляции:

$$C_{f_1, f_2}(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1(x, y) f_2(x - \Delta x, y - \Delta y). \quad (2.11)$$

Несложно убедиться, что положение корреляционного максимума при линейном глобальном преобразовании яркости  $\alpha f_1(x, y) + \beta$  не меняется (только при  $\alpha < 0$  максимум превращается в минимум). Таким образом, данный подход позволяет сопоставлять изображения при линейных глобальных искажениях яркости.

Если смещения  $\Delta x$  и  $\Delta y$  пробегает значения от  $-N$  до  $N$ , то несложно убедиться, что число операций при вычислении кросскорреляционного поля по формулам (2.10) или (2.11) будет пропорционально  $N^4$ .

Однако существует возможность построения кросскорреляционного поля  $C_{f_1, f_2}(\Delta x, \Delta y)$  за меньшее число операций, если воспользоваться быстрым преобразованием Фурье и тем фактом, что Фурье-образ кросскорреляционного поля может быть получен как произведение спектров изображений.

По определению, прямое и обратное дискретные преобразования Фурье, ДПФ, (см. рис. 2.10) вычисляются как

$$\begin{aligned} \tilde{f}(u, v) &= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-i \frac{2\pi}{N}(ux+vy)}, \\ f(x, y) &= \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \tilde{f}(u, v) e^{i \frac{2\pi}{N}(ux+vy)}. \end{aligned} \quad (2.12)$$

где  $u$  и  $v$  – пространственные частоты. При вычислении пространственного спектра по этим формулам число операций также пропорционально  $N^4$ , однако алгоритм быстрого преобразования Фурье позволяет произвести вычисление за время, пропорциональное  $N^2 \ln N$ .

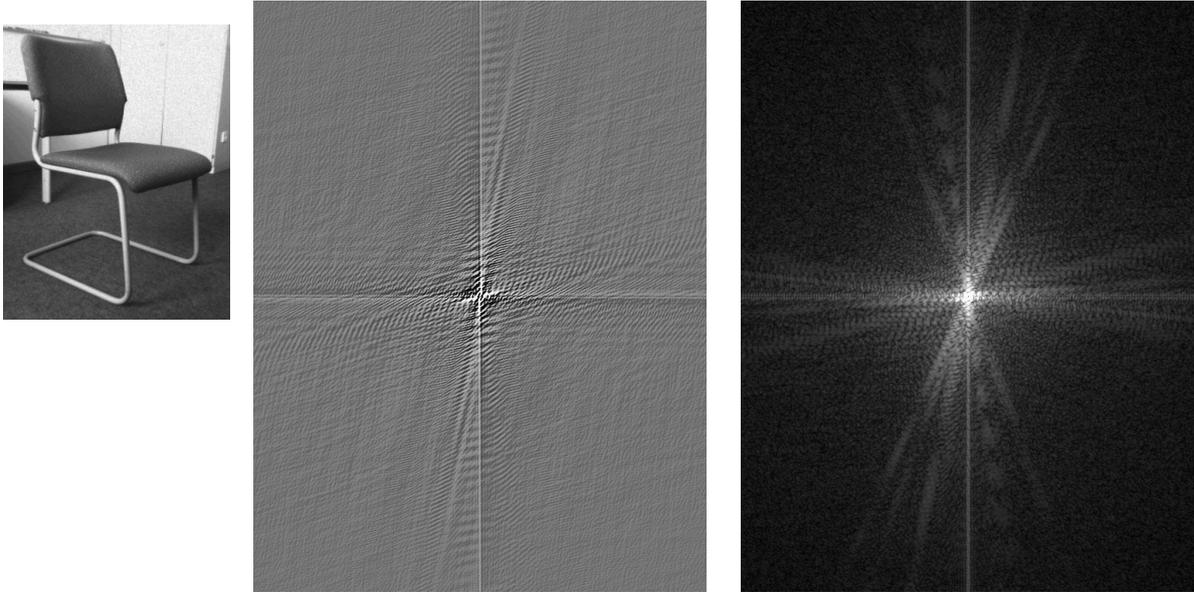


Рис. 2.10. Пример вычисления преобразования Фурье: (а) исходное изображение; (б) вещественная компонента спектра; (в) амплитудный (энергетический) спектр

Кросскорреляционное поле тогда может быть получено как

$$\begin{aligned}
 C_{f_1, f_2}(\Delta x, \Delta y) &= \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \tilde{f}_1(u, v) \tilde{f}_2^*(u, v) e^{i \frac{2\pi}{N}(u\Delta x + v\Delta y)} = \\
 &= \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1(x, y) f_2^*(x - \Delta x, y - \Delta y),
 \end{aligned} \tag{2.13}$$

что также может быть вычислено за время, пропорциональное  $N^2 \ln N$  ( $f^*$  – комплексное сопряжение). Таким образом, общее время работы при вычислении кросскорреляционного поля через ДПФ будет пропорционально  $N^2 \ln N$ , хотя коэффициент пропорциональности значительно больше, чем при непосредственном вычислении по формуле (2.11), в связи с чем данный подход нерационально использовать при малых  $N$ .

Необходимо отметить, что в алгоритмах ДПФ изображение трактуется как периодический сигнал, то есть копии изображения как бы приставлены друг к другу, из-за чего появляется резкий переход от одного края изображения к другому. Это вызывает появление «креста» в центре спектра изображения (см. рис. 2.10) и, как мы увидим ниже, ухудшение корреляционного отклика. Плавное приведение интенсивностей у краев к среднему значению устраняет эффект края. Как видно из рис. 2.11, центральный «крест», который отчетливо виден на рис. 2.10, пропадает. При этом также рекомендуется расширять изображение, помещая его на поле однородной яркости (тогда коррелируемое с ним изображение не будет попадать на соседние периоды).

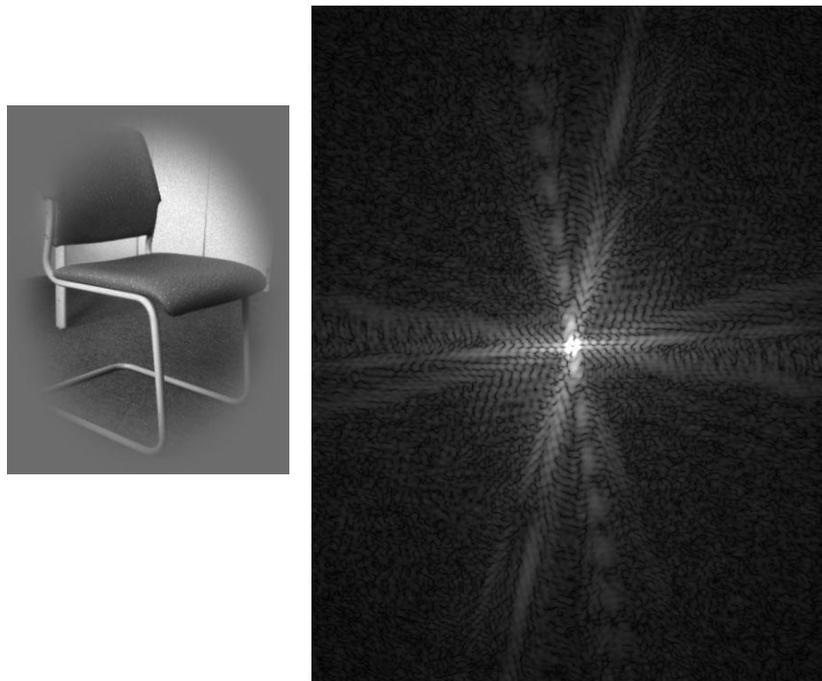


Рис. 2.11. Изображение, усредненное к краям, и его амплитудный спектр

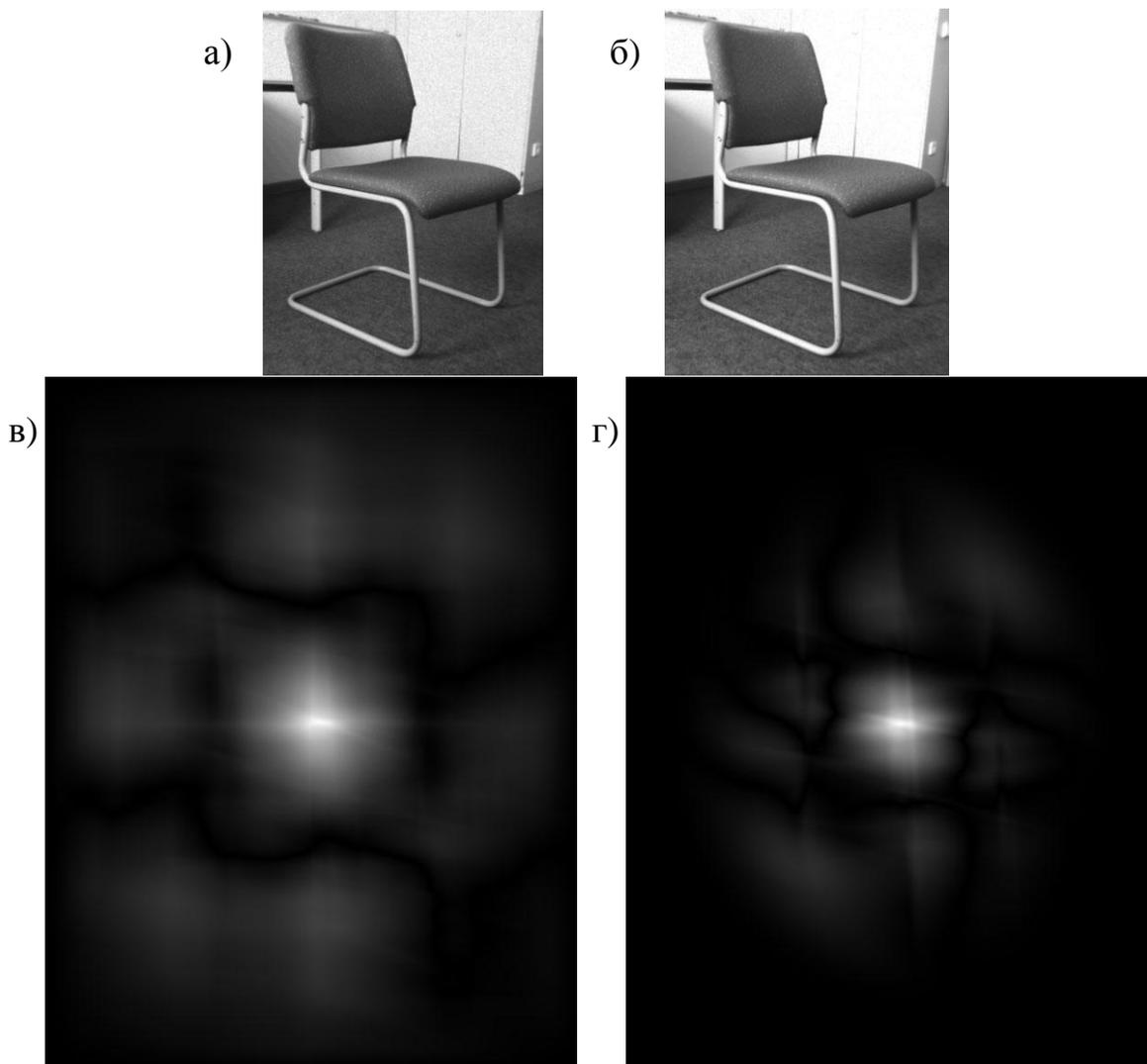


Рис. 2.12. (а) и (б) сопоставляемые изображения, снятые с разных ракурсов; (в) кросскорреляционное поле двух изображений, вычисленное с помощью ДПФ; (г) кросскорреляционное поле изображений со сглаженными краями

На рис. 2.12 представлен пример вычисления кросскорреляционного поля без устранения эффекта края и с его устранением. На рисунке приведены два изображения, полученные с разных ракурсов. Поскольку это изображения трехмерного объекта, а проекция центральная, то геометрическое преобразование изображения не описывается одним общим смещением (но сдвиги для всех точек достаточно близки), в связи с чем максимум на кросскорреляционном поле не выражен четко. Однако видно, что максимум может быть локализован, причем при использовании сглаживания к краю кросскорреляционное поле выглядит контрастнее, и максимум может быть определен точнее. В действительности, эффект края приводит к появлению ложного максимума в центре поля (точке, соответствующей смещению  $(0, 0)$ ) и к размытию максимума, соответствующего истинному смещению.

Кроме того, из рис. 2.12 видно, что значение максимума на кросскорреляционном поле не сильно отличается от его окружения. Существуют методы, позволяющие осуществлять обнаружение и локализацию максимума с более высокой точностью.

### Фазовая корреляция

Корреляционные методы, описанные выше, называются амплитудно-фазовыми, поскольку в них используется как амплитудная, так и фазовая спектральная информация. Помимо этого используется фазовая корреляция, при вычислении которой амплитудная информация не используется:

$$C_{f_1, f_2}^{\phi}(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \frac{\tilde{f}_1(u, v)}{|\tilde{f}_1(u, v)|} \cdot \frac{\tilde{f}_2^*(u, v)}{|\tilde{f}_2^*(u, v)|} e^{i \frac{2\pi}{N}(u\Delta x + v\Delta y)}. \quad (2.14)$$

Напомним, что Фурье-образы  $\tilde{f}_1(u, v)$  и  $\tilde{f}_2(u, v)$  являются комплекснозначными, даже если исходные изображения представлены вещественнозначными функциями. Значения  $|\tilde{f}_1(u, v)|$  и  $|\tilde{f}_2(u, v)|$  являются модулями соответствующих комплексных величин.

Особенность фазовой корреляции заключается в том, что для нее автокорреляционное поле некоторой функции представляет собой дельта-функцию, в чем несложно убедиться

$$C_{f, f}^{\phi}(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} e^{i \frac{2\pi}{N}(u\Delta x + v\Delta y)} = \begin{cases} 1, & (\Delta x, \Delta y) = (0, 0) \\ 0, & (\Delta x, \Delta y) \neq (0, 0) \end{cases}. \quad (2.15)$$

На рис. 2.13 представлено кросскорреляционное поле (в случае фазовой корреляции) изображений, приведенных на рис. 2.12а,б. Видно, что здесь максимум может быть локализован с большей точностью (при этом видно, что в данном случае выделяется несколько максимумов, соответствующих сдвигу объекта на переднем плане и фона, имеющих разные смещения).

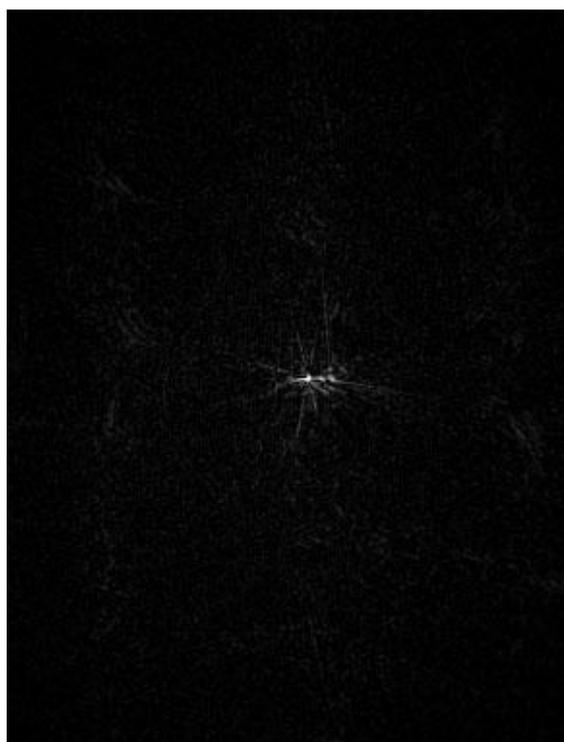


Рис. 2.13. Пример кросскорреляционного поля, получающийся в случае фазовой корреляции вместо поля, представленного на рис. 2.12г.

Следует отметить, что в случае фазовой корреляции также полезным является сглаживание изображений к краю с дальнейшим расширением области изображений с заполнением области расширения средним значением (см. рис. 2.14).

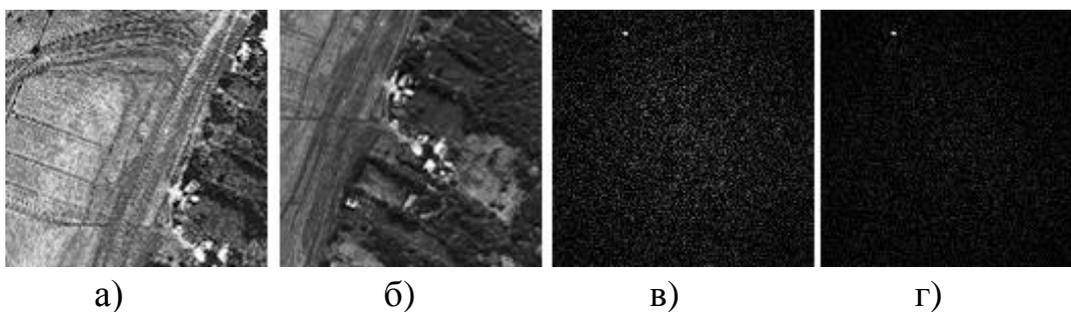


Рис. 2.14. Пример уменьшения шумов на кросскорреляционном поле в случае фазовой корреляции при устранении эффекта края: а) и б) исходные изображения; в) кросскорреляционное при использовании фазовой корреляции; г) кросскорреляционное поле при устранении эффекта края

Кроме лучшей локализации максимума метод фазовой корреляции в ряде случаев может также оказаться несколько менее чувствительным к нелинейным или локальным преобразованиям яркости, чем метод амплитудно-фазовой корреляции. Однако в этом подходе не используется большой объем информации, заключенный в амплитудной составляющей спектра, что может привести в ряде случаев и к понижению вероятности обнаружения максимума, соответствующего истинному смещению. Общим недостатком описанных методов является сильное ограничение на

допустимое пространственное преобразование изображений, а именно, оно должно являться преобразованием сдвига.

### *Метод Фурье-Меллина*

При вычислении корреляционного поля по формуле (2.11) нет принципиальных ограничений на то, чтобы варьировать только взаимное смещение двух изображений. В равной степени могут перебираться не только все возможные сдвиги, но также и углы поворота или, к примеру, взаимные масштабы изображений. Когда выбран конкретный набор параметров пространственного преобразования, после его выполнения может быть посчитана величина корреляции между одним изображением и преобразованным другим изображением. По максимуму корреляционного отклика можно таким образом определить параметры пространственного преобразования (при условии, что яркостные искажения не слишком отличаются от линейных).

При этом, однако, возникает трудность, связанная с вычислительной сложностью алгоритма сопоставления. Если полагать, что каждый параметр преобразования может принимать одно из  $N$  значений, то необходимое число операций будет пропорционально  $N^{m+2}$ , где  $m$  – число параметров. Поскольку уже для преобразования сдвига (при  $m=2$ ) оказывается критичным повышение быстродействия алгоритма за счет использования БПФ, то при большем числе параметров время сопоставления оказывается неприемлемым.

Однако вместо поиска максимума во всем пространстве параметров взаимного геометрического преобразования изображений существует возможность так преобразовать изображения, чтобы результат преобразования оказался инвариантным к некоторой подгруппе исходной группы геометрического преобразования.

В частности, амплитудный спектр изображения  $|\tilde{f}_1(u, v)|$  остается постоянным при сдвиге этого изображения (меняется лишь фаза гармоник пространственного спектра, но не их амплитуда). Поскольку помимо взаимных сдвигов изображения одного и того же объекта часто подвержены вращению и масштабированию, возникает вопрос, как преобразуется спектр при масштабировании и вращении изображения. Несложно убедиться, что масштабирование и вращение исходного изображения приводит к такому же масштабированию и вращению его спектра. Таким образом, амплитудный спектр изображения  $|\tilde{f}_1(u, v)|$  инвариантен к сдвигу изображения, но сохраняет тот же масштаб и угол поворота.

Можно было бы рассматривать амплитудные спектры  $|\tilde{f}_1(u, v)|$  и  $|\tilde{f}_2(u, v)|$  в качестве изображений, которые необходимо сопоставить при условии, что они отличаются только масштабом и углом поворота. При этом, казалось бы, нужно перебирать разные значения этих параметров пространственного преобразования и вычислять значение корреляции по

формуле (2.11). Однако и здесь можно воспользоваться быстрым вычислением корреляционного поля через БПФ, если перейти в такую систему координат, в которой масштаб и поворот превращаются в сдвиги по двум осям. Такой системой координат является полярно-логарифмическая система координат, переход к которой определяется соотношениями:

$$\varphi = \arctan(y/x), \rho = \ln \sqrt{x^2 + y^2} . \quad (2.16)$$

На рис. 2.15 представлен пример перехода к полярно-логарифмическим координатам для изображений, отличающихся масштабом и углом поворота. Из рисунка видно, что изменение масштаба и угла поворота действительно приводит к сдвигу изображения в данной системе координат (напомним, однако, что данное преобразование координат подразумевается применять не к исходным изображениям, которые могут быть смещены друг относительно друга, а их амплитудным спектрам).

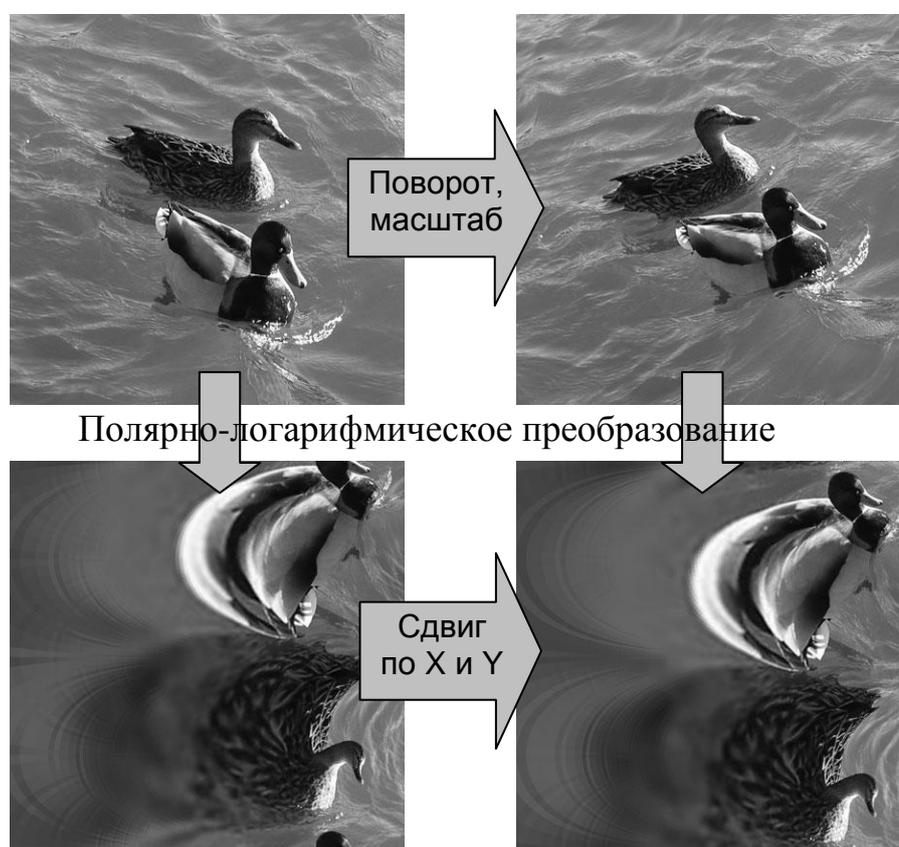


Рис. 2.15. Пример перехода в полярно-логарифмическую систему координат при масштабировании и вращении изображения

В итоге, после отображения амплитудного спектра изображения в полярно-логарифмическую систему координат значения взаимного масштаба и вращения могут быть найдены как сдвиги, для чего может быть использован обычный корреляционный метод с вычислением кросскорреляционного поля с помощью БПФ. После нахождения масштаба и угла поворота они могут быть компенсированы для исходных изображений, после чего уже может быть найден и взаимный сдвиг.

Таким образом, мы приходим к методу Фурье-Меллина, позволяющему сопоставлять изображения при взаимном преобразовании из группы подобия за время порядка  $N^2 \ln N$ . Этот метод состоит из следующих шагов

1. Для данных изображений  $f_1$  и  $f_2$  построить их амплитудные спектры  $|\tilde{f}_1(u, v)|$  и  $|\tilde{f}_2(u, v)|$ .
2. Перевести амплитудные спектры в полярно-логарифмическую систему координат:
 
$$|\tilde{f}^{(pl)}(\rho, \varphi)| = |\tilde{f}(e^\rho \cos(\varphi), e^\rho \sin(\varphi))|.$$
3. Построить кросскорреляционное поле «изображений»  $|\tilde{f}_1^{(pl)}(\rho, \varphi)|$  и  $|\tilde{f}_2^{(pl)}(\rho, \varphi)|$  в плоскости параметров  $(\rho, \varphi)$  и найти на нем максимум  $(\rho_m, \varphi_m)$ .
4. Компенсировать взаимный масштаб  $e^{\rho_m}$  и угол вращения  $\varphi_m$  для исходных изображений, после чего построить кросскорреляционное поле и найти на нем максимум  $(\Delta x_m, \Delta y_m)$ , который будет соответствовать взаимному сдвигу.

Таким образом, будут найдены все параметры преобразования группы подобия.

На рис. 2.16 представлен пример успешного совмещения пары изображений методом Фурье-Меллина. Как видно из рисунка, изображения отличаются не только сдвигом, но и углом поворота.



Рис. 2.16. пример совмещения изображений методом Фурье-Меллина

Хотя метод Фурье-Меллина позволяет расширить класс допустимых взаимных пространственных преобразований, при этом несколько снижается надежность определения параметров преобразования, поскольку на каждом шаге данного метода происходит потеря части информации (при построении амплитудного спектра причина этой потери очевидна, однако заметные потери появляются и при переходе к полярно-логарифмической системе координат из-за дискретности координатной сетки).

Несмотря на расширение класса допустимых пространственных преобразований в методе Фурье-Меллина, это решение оказывается недостаточно общим, поскольку оно не подходит как для аффинных или проективных глобальных преобразований, так и для локальных преобразований, которые часто встречаются в реальных задачах компьютерного зрения. В рамках корреляционного подхода существует возможность дальнейшего расширения класса допустимых пространственных преобразований.

#### *Метод локальной корреляции*

Многие методы сопоставления изображений работают с ограниченными моделями пространственных преобразований (например, группы подобия или аффинной группы). При этом преобразования выбранного являются хотя и не точной, но неплохой аппроксимацией реального взаимного преобразования изображений. В результате, эти методы оказываются относительно работоспособными, но не точными. После компенсации основных компонентов пространственного преобразования остаются некомпенсированными компоненты следующего порядка малости.

К примеру, изображения, представленные на рис. 2.12а,б обладают некоторым взаимным смещением, после компенсации которого остаются небольшие собственные смещения для каждой точки изображений. Также, например, и для некоторых аэрокосмических изображений после компенсации аффинного преобразования может остаться незначительная проективная компонента, которая может быть описана в терминах локальных сдвигов, приписанных каждому пикселю.

Сходная ситуация имеет место для изображений биологических объектов, например, изображений, полученных в результате томографии головного мозга. В целом изображения могут быть пространственно совмещены (если компенсирован поворот головы относительно томографа), но отдельные области мозга могут быть немного сдвинуты друг относительно друга.

Если глобальное взаимное преобразование компенсировано, то окрестности сопряженных точек двух изображений должны отличаться незначительно. Это приводит к общей идее использовать корреляцию не для всей области изображений, а только для их фрагментов, взятых вокруг некоторых точек (см. рис. 2.17), что приводит к возможности построения поля локальных смещений (см. рис. 2.18).

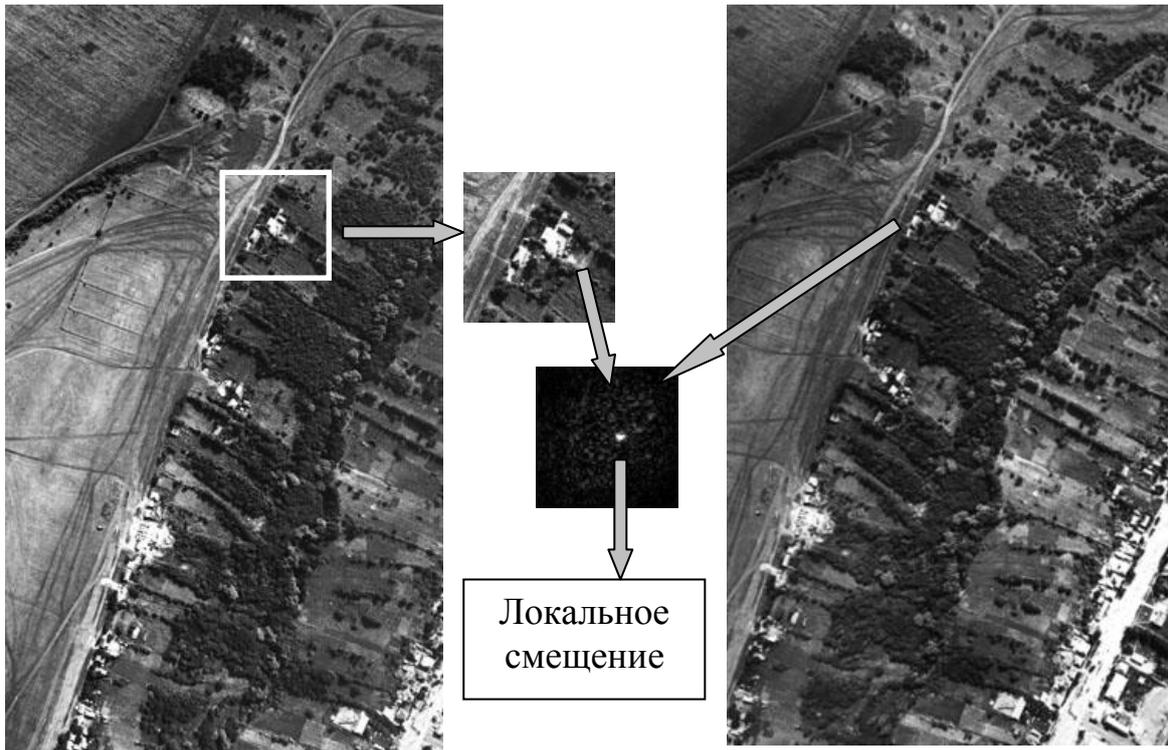


Рис. 2.17. Пример вычисления локального смещения по фрагменту изображения в рамках корреляционного подхода



Рис. 2.18. Пример построения поля локальных смещений путем выполнения корреляции фрагментов, взятых вокруг набора точек, расположенных на регулярной сетке (белой линией отмечен овраг, по разные стороны от которого местность имеет разный наклон, что вызывает смещение соответствующих точек на изображениях, снятых с разных ракурсов)

Существует много вариантов реализации идеи локальной корреляции, однако мы не будем углубляться в технические подробности. Рассмотрим лишь пример использования здесь подхода с переменным разрешением. Поскольку при локальной корреляции берутся фрагменты фиксированного размера вокруг точек с одинаковыми координатами, то при локальном смещении, большем размера окна, локальное смещение в принципе не может быть определено. Однако если поле локальных смещений строить для изображений с пониженным разрешением, а далее их уточнять с постепенным повышением разрешения, то ограничение на допустимые локальные сдвиги будут слабее (рис. 2.19).

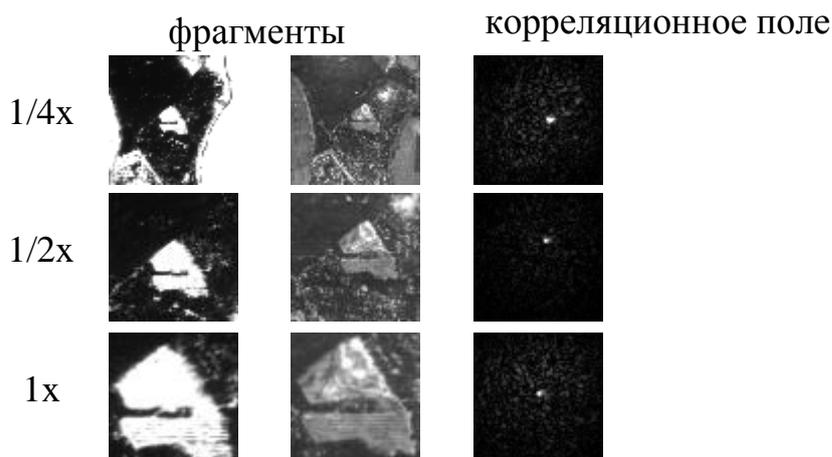


Рис. 2.19. Пример уточнения локального смещения для некоторой пары изображений

Подход с переменным разрешением позволяет расширить допустимые локальные сдвиги. Эти сдвиги могут появляться вследствие неполной компенсации взаимного глобального пространственного преобразования изображений. Однако если не полностью компенсированными остаются взаимный масштаб и поворот, то локальные фрагменты, взятые даже вокруг сопряженных точек (то есть с относительным нулевым смещением), будут подвержены тому же взаимному масштабу и повороту. Метод локальной корреляции, как и глобальный корреляционный метод, работающий только со сдвигами, весьма чувствителен к подобным геометрическим искажениям (см. рис. 2.20).

Как отмечалось ранее, корреляционные методы очень чувствительны к преобразованиям яркости, связанным с изменением условий освещения, сменой типов сенсоров и собственной изменчивостью объектов (например, сезонно-суточными изменениями для аэрокосмических изображений). Метод локальной корреляции менее чувствителен к подобным яркостным искажениям, поскольку допускает линейное преобразование яркости в каждом отдельном фрагменте изображения, а не только единое линейное преобразование яркости для всего изображения. Однако остается проблема компенсации глобального пространственного преобразования в условиях сложных яркостных искажениях.

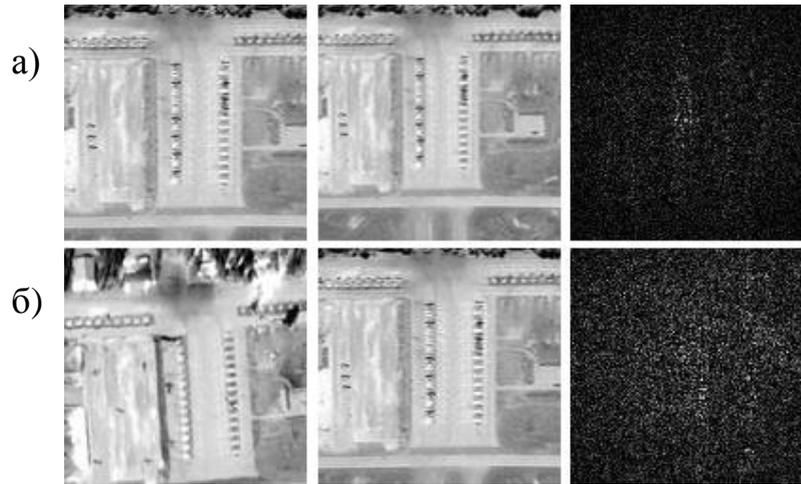


Рис. 2.20. Пример разрушения корреляционного отклика при фазовой корреляции (а) фрагмента со своей копией, уменьшенной в 1.1 раза; (б) фрагментов двух разных изображений одной и той же местности при некомпенсированном угле поворота, равном  $3^\circ$

### *Совмещение изображений путем максимизации взаимной информации*

Можно считать, что корреляционные методы опираются на функциональные представления изображений. Можно обратиться и к более общим стохастическим моделям изображений, в которых яркости пикселей трактуются как отсчеты некоторой случайной величины. Тогда вероятности случайных величин, соответствующих двум совмещаемым изображениям, не будут статистически независимыми.

В случае использования простейшей стохастической модели яркости во всех точках изображения считаются реализациями независимых и одинаково распределенных случайных величин. Тогда для изображения  $f(x)$  как реализации некоторой случайной функции  $F$  может быть вычислена его энтропия

$$H(F) = E_F [I(f)] = \sum_{f \in V} I(f) P(f) = - \sum_{f \in V} P(f) \log_2 P(f) = - \sum_{\mathbf{x} \in G} \log_2 P(f(\mathbf{x})),$$

где  $G$  и  $V$  – области определения и значения  $f(\mathbf{x})$ ,  $E$  – математическое ожидание случайной величины,  $I$  – количество информации,  $P$  – вероятность. В данной формуле  $P(f)$  можно считать гистограммой изображения  $f(\mathbf{x})$ .

Условная энтропия двух изображений  $F_1$  и  $F_2$  примет форму

$$\begin{aligned} H(F_1 | F_2) &= - \sum_{f_2} \left( \sum_{f_1} P(f_1 | f_2) \log_2 P(f_1 | f_2) \right) P(f_2) = \\ &= - \sum_{f_2} \sum_{f_1} P(f_1, f_2) \log_2 P(f_1 | f_2) \end{aligned}$$

Взаимная информация, содержащаяся в изображениях, определяется как

$$I(F_1, F_2) = E_{F_1 F_2} [I(f_1; f_2)] = \sum_{f_1} \sum_{f_2} P(f_1, f_2) \log_2 \frac{P(f_1, f_2)}{P(f_1)P(f_2)}. \quad (2.17)$$

Для взаимной информации выполняются соотношения

$$I(F_1, F_2) = H(F_1) - H(F_1 | F_2) = H(F_2) + H(F_1) - H(F_1, F_2) = \\ = H(F_2) - H(F_2 | F_1) = I(F_2, F_1). \quad (2.18)$$

Кроме того, основным свойством взаимной информации является то, что ее значение для двух случайных величин инвариантно к произвольным функциональным преобразованиям этих величин и изменяется только при изменении степени их статистической независимости.

Иными словами, равенство нулю взаимной информации означает, что для любых невырожденных функций  $\varphi_1$  и  $\varphi_2$  выполняется равенство

$$E_{F_1 F_2} [\varphi_1(f_1) \varphi_2(f_2)] - E_{F_1} [\varphi_1(f_1)] E_{F_2} [\varphi_2(f_2)] = 0. \quad (2.19)$$

В то же время, равенство нулю коэффициента корреляции означает, что это равенство выполняется лишь для линейных преобразований  $\varphi_1$  и  $\varphi_2$ . Таким образом, можно считать (в контексте сопоставления изображений), что критерий взаимной информации расширяет корреляционный критерий на случай нелинейных преобразований яркости (см. рис. 2.21).

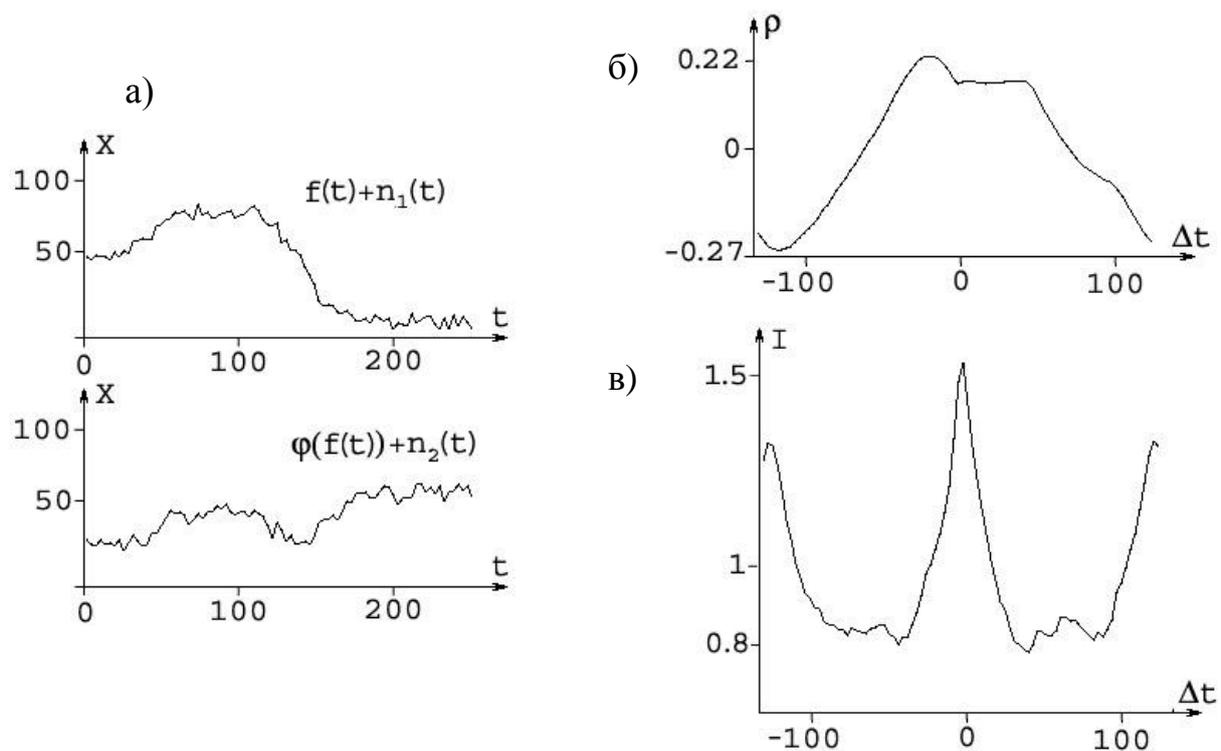


Рис. 2.21: а) срезы двух изображений (с нулевым взаимным сдвигом) при нелинейных искажениях яркости; б) результат вычисления корреляции (максимум смещен); в) результат вычисления взаимной информации при разных сдвигах (выраженный максимум соответствует истинному смещению)

Таким образом, вместо корреляционных методов могут использоваться методы совмещения изображений путем максимизации взаимной информации, в которых перебираются разные значения параметров пространственного преобразования, при которых вычисляется значение взаимной информации и находится максимум.

Несмотря на возможность работать с нелинейными преобразованиями яркости, в методах совмещения путем максимизации взаимной информации

остается ограничение на то, что это преобразование должно быть глобальным (то есть единообразно применимым к любой точке изображения). Это в определенной степени верно, например, для случая смены сенсора, однако изменение условий освещения и собственная изменчивость объектов могут вызывать локальные, неоднородные, изменения яркости. То есть и здесь проблема инвариантности яркостным преобразованиям не решается. Кроме того, для методов совмещения изображений путем максимизации взаимной информации не существует хороших способов оптимизации вычислений, как для корреляционных методов.

Итак, несмотря на все улучшения, методы, использующие низкоуровневые представления, не предоставляют универсального средства решения проблемы сопоставления изображений. Приемлемые решения имеются только для преобразований группы подобия, а уже для аффинных преобразований методы оказываются весьма ресурсоемкими и при этом неинвариантными по отношению к сложным преобразованиям яркости. В связи с этим низкоуровневые методы сопоставления имеют ограниченную сферу применения, и оказывается необходимым при решении задачи сопоставления изображений привлекать более высокоуровневые представления.

### Вопросы и упражнения

1. Докажите, что положение максимума на корреляционном поле не меняется при линейных преобразованиях яркости исходных изображений.
2. Вывести формулу для автокорреляционного поля (при фазовой корреляции) произвольной функции в непрерывном и дискретном случаях.
3. Для каких целей в корреляционных методах производится расширение изображений и приведение их яркости к среднему значению у края?
4. Докажите, что вращение и масштабирование исходного изображения приводит к вращению и масштабированию его спектра (в непрерывном случае).
5. Обоснуйте утверждение, согласно которому метод Фурье-Меллина реализуется за время, пропорциональное  $N^2 \ln N$ , где  $N$  – линейный размер изображения.
6. Из каких шагов состоит метод Фурье-Меллина? Какие операции над изображениями при этом используются и для каких целей?
7. В чем заключается основная идея метода локальной корреляции?
8. Какова вычислительная сложность метода совмещения изображений путем максимизации взаимной информации в случае преобразования сдвига?

9. Каковы основные ограничения низкоуровневых методов сопоставления изображений?
10. Какие низкоуровневые методы вы знаете, и какие пространственные преобразования изображений являются для них допустимыми?

### 2.3. Контурные и структурные методы сопоставления изображений

#### *Преобразование расстояния*

Низкоуровневые методы сопоставления изображений сталкиваются с рядом упомянутых выше проблем, среди которых неинвариантность к различным видам изменчивости изображений, ограничения на допустимые пространственные преобразования, а также достаточно большая вычислительная сложность.

В целях сопоставления изображений естественно использовать представления изображений более высоких уровней, в которых извлекается инвариантная информация и уменьшается объем нерелевантных данных. В действительности, тип привлекаемого представления изображений и является одной из основных характеристик конкретного метода их сопоставления.

Вслед за низкоуровневыми представлениями следующими по уровню абстрактности идут контурные представления. Способы построения контуров весьма разнообразны, однако их дальнейшее использование при сопоставлении изображений примерно одинаково, поэтому при разработке метода сопоставления на основе контуров, последние рассматриваются в качестве исходной информации вместо изображений.

Пусть  $\mathbf{C}_1 = \{(x_{1,i}, y_{1,i})\}_{i=1}^{n_1}$  и  $\mathbf{C}_2 = \{(x_{2,i}, y_{2,i})\}_{i=1}^{n_2}$  – совокупности контурных точек на первом и втором изображениях соответственно. Как правило, значения  $n_1$  и  $n_2$  достаточно велики, хотя и существенно меньше, чем общее число точек в изображении (обычно число контурных точек в несколько раз больше, чем линейный размер соответствующего изображения). Кроме того, контурные точки неуникальны, в том смысле, что одну контурную точку, взятую в отдельности, нельзя отличить от другой контурной точки. Это приводит к тому, что непосредственное отождествление контурных точек невозможно.

В связи с этим, в контурных методах сопоставления изображений, как и в корреляционных методах, производится поиск решения в пространстве параметров взаимного геометрического преобразования изображений. Иными словами, если это преобразование задается как пара функций  $\varphi_x(x, y, \mathbf{w})$  и  $\varphi_y(x, y, \mathbf{w})$  с неизвестным вектором параметров  $\mathbf{w}$ , то необходимо перебирать различные значения этих параметров, оптимизируя некоторую меру сходства двух наборов контуров  $\{\varphi_x(x_{1,i}, y_{1,i}, \mathbf{w}), \varphi_y(x_{1,i}, y_{1,i}, \mathbf{w})\}_{i=1}^{n_1}$  и  $\{(x_{2,i}, y_{2,i})\}_{i=1}^{n_2}$ .

Естественной мерой сходства является число совпадающих точек в двух наборах. Если параметры пространственного преобразования выбраны неправильно, то число совпадающих точек должно быть мало, и, напротив, при правильных значениях параметров число совпадающих точек должно быть максимально. Однако требование точного совпадения является слишком жестким. В этом несложно убедиться, если попытаться корреляционно сопоставить контурные изображения (см. рис. 2.22). Как видно из рисунка, корреляционное поле «распадается» на сумму компонентов, связанных с каждым из контуров.

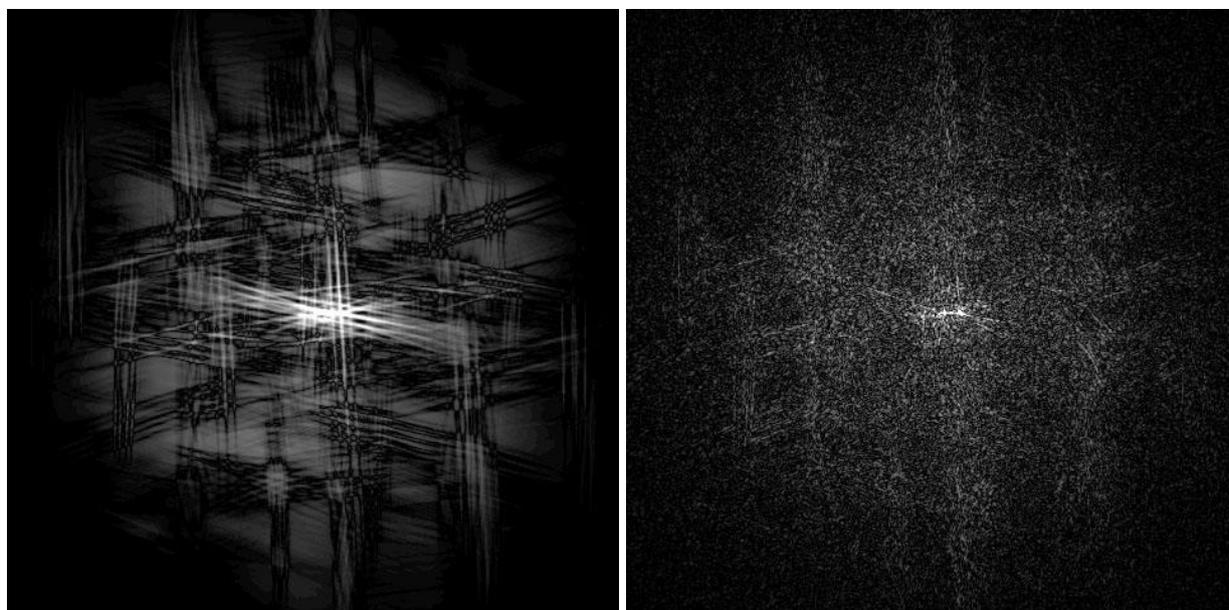
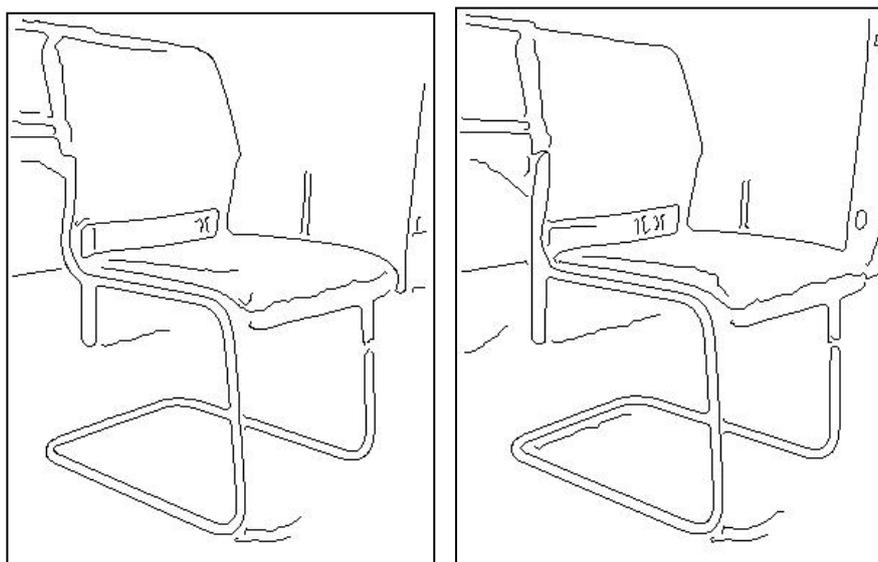


Рис. 2.22. Контурные изображения, выделенные на изображениях, представленных на рис. 2.12а и 2.12б и результаты вычисления кросскорреляционного поля в случае амплитудно-фазовой и фазовой корреляции

В связи с этим следует искать не точно совпадающую, а ближайшую контурную точку, и расстояние  $d$  до нее определяет качество соответствия, которое можно задать как  $s/(s + d)$ , где параметр  $s$  задает отклонение, на

котором значение меры сходства падает в два раза (своего рода "толщина" контуров). Однако непосредственный поиск ближайшей точки потребует  $n_2$  операций для каждой контурной точки первого изображения и для каждого значения параметров преобразования  $w$ .

Для сокращения числа операций выполняется т.н. *преобразование расстояния*, которое в случае использования евклидовой метрики для произвольного множества  $C$  можно определить следующим образом

$$D_C(x, y) = \min_{(\chi, \upsilon) \in C} \left( (x - \chi)^2 + (y - \upsilon)^2 \right)^{0.5}. \quad (2.20)$$

Результат действия преобразования расстояния контуров можно рассматривать в качестве изображения (см. рис. 2.23).

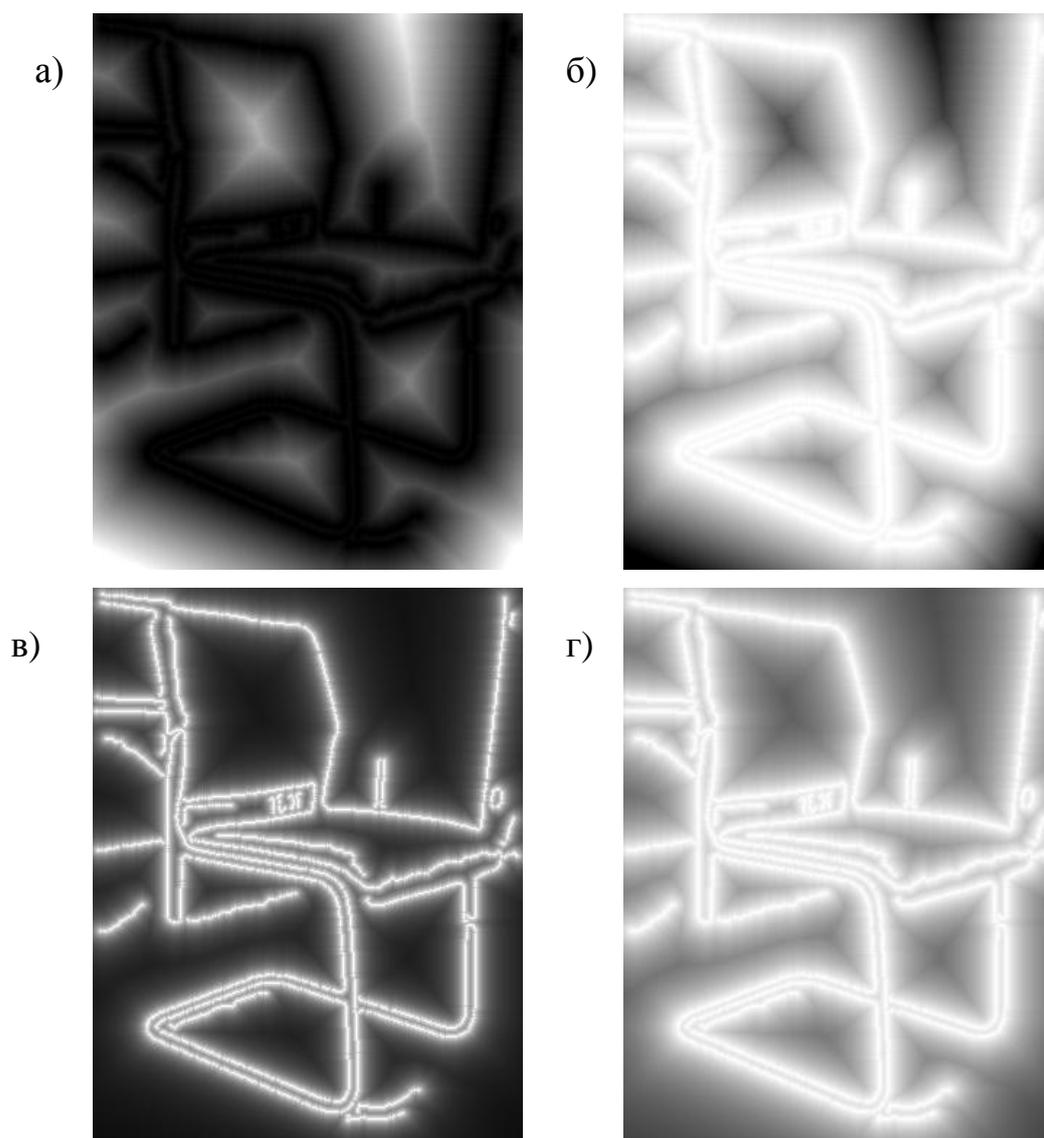


Рис. 2.23. Результат вычисления преобразования расстояния (а) и его инверсии (б) по контурному изображению на рис. 2.22, а также преобразования  $s/(s+D_c)$  при  $s=5$  и  $s=30$  для (в) и (г) соответственно

Преобразование расстояния позволяет вычислить *меру Хаусдорфа* для двух множеств. Эта мера характеризует расстояние между множествами и определяется как

$$h(\mathbf{C}_1, \mathbf{C}_2) = \max_{(x,y) \in \mathbf{C}_1} D_{\mathbf{C}_2}(x, y) = \max_{(x,y) \in \mathbf{C}_1} \min_{(\chi, \upsilon) \in \mathbf{C}_2} \left( (x - \chi)^2 + (y - \upsilon)^2 \right)^{0.5}. \quad (2.21)$$

Может использоваться также частичная мера Хаусдорфа, в которой поиск максимума осуществляется не по всем точкам множества  $\mathbf{C}_1$ , а лишь по некоторому количеству ближайших к множеству  $\mathbf{C}_2$  точек.

Поскольку пространственное преобразование применяется только к контурам одного изображения, контуры опорного изображения остаются неизменными, а значит, неизменным является и результат применения к ним преобразования расстояния, которое достаточно вычислить в процессе сопоставления лишь один раз.

Кроме того, существует быстрый алгоритм вычисления этого преобразования, требующий лишь порядка  $N^2$  операций, для изображения  $N \times N$  пикселей (данный алгоритм полностью эквивалентен алгоритму поиска кратчайшего пути на графе).

Таким образом, критерий качества соответствия изображений при их контурном сопоставлении (этот вариант контурного сопоставления называется *чамферным* сопоставлением от англ. *chamfer* – выемка, бороздка) выражается в форме

$$Q(\mathbf{w}) = \frac{1}{n_1} \sum_{(x,y) \in \mathbf{C}_1} \frac{s}{s + D_{\mathbf{C}_2}(\varphi_x(x, y, \mathbf{w}), \varphi_y(x, y, \mathbf{w}))}. \quad (2.22)$$

Величину  $Q(\mathbf{w})$  необходимо максимизировать по вектору параметров пространственного преобразования  $\mathbf{w}$ . Отметим, что здесь, по существу, порядок контурных точек во множествах  $\mathbf{C}_1$  и  $\mathbf{C}_2$  не важен, то есть могут использоваться не только контуры, но и просто наборы граничных точек, найденных в рамках признакового подхода.

Под знаком суммы может стоять и другая функция от преобразования расстояния. Однако, как правило, изменение вида функции не сильно сказывается на результате, поскольку основным параметром в данном случае выступает ширина «бороздки» вокруг контура, задаваемая параметром  $s$ . В ряде случаев может производиться непосредственное суммирование значений преобразования расстояния (и сумму эту необходимо минимизировать)

$$Q_2(\mathbf{w}) = \frac{1}{n_1} \sum_{(x,y) \in \mathbf{C}_1} D_{\mathbf{C}_2}(\varphi_x(x, y, \mathbf{w}), \varphi_y(x, y, \mathbf{w})) \quad (2.23)$$

или использование меры Хаусдорфа.

Здесь, однако, нужно иметь в виду, что если на втором изображении присутствует частично загороженный объект, то на первом изображении будут присутствовать контуры частей объекта, отсутствующие на втором изображении, и значение  $D_{\mathbf{C}_2}$  будет весьма велико, а критерий  $Q_2(\mathbf{w})$  будет иметь высокое значение, свидетельствующее о плохом сопоставлении, даже для правильных значений параметров пространственного преобразования. Иными словами, тогда как в критерии  $Q(\mathbf{w})$  считается число правильно

сопоставленных контурных точек, в критерии  $Q_2(\mathbf{w})$  считается число ошибок. Выбор критерия достаточно произволен и должен осуществляться эмпирически для конкретной задачи.

Методы чамферного сопоставления достаточно просты и заключаются в вычислении критерия  $Q(\mathbf{w})$  для разных значений параметров  $\mathbf{w}$ . При этом одно вычисление  $Q(\mathbf{w})$  требует порядка  $n_1$  операций ( $n_1 \propto N$ ). На рис. 2.24. представлен пример вычисления этого критерия для случая, когда пространственно преобразование является преобразованием сдвига (размерность вектора параметров равна двум).



Рис. 2.24. Значение меры сходства  $Q(\mathbf{w})$  двух контурных изображений, представленных на рис. 2.22, при различных смещениях; положение максимума соответствует среднему смещению изображений

К сожалению, уже для преобразования подобия число операций, требуемых для чамферного сопоставления, составляет порядка  $N^4$ , и может оказаться весьма большим. Здесь, однако, могут использоваться разные способы ускорения поиска. Во-первых, точки множества  $S_1$  могут прореживаться, чтобы сократить число действий при суммировании (2.22); при этом преобразование расстояния вычисляется один раз по всем точкам множества  $S_2$ .

Кроме того, может привлекаться подход с переменным разрешением, в котором сначала решение находится на уменьшенных в несколько раз изображениях, а затем постепенно уточняется (при этом пространство поиска сокращается). Можно также сам поиск производить иерархически на исходном разрешении: перебирать все значения вектора параметров  $\mathbf{w}$  на крупной сетке (например, перебирать смещения с шагом 5 пикселей), а затем уточнять положение найденного максимума.

Отметим также, что если контуры разделены на связные цепочки, то для каждой цепочки чамферное сопоставление может вестись независимым образом. В результате могут быть измерены не только параметры глобального пространственного преобразования, но также и локальные смещения.

Одной из модификаций, о которой необходимо обязательно упомянуть, является использование при чамферном сопоставлении не только информации о положении точек контуров, но и о направлении градиента в

этих точках (который, как правило, перпендикулярен к касательной в точке контура). Поскольку контурные точки обычно строятся как точки локальных максимумов на градиентном поле изображения, то с этими точками можно связать определенную ориентацию (см. рис. 2.25). Дополнительный учет расхождения ориентаций сопоставляемых точек в уравнении для меры  $Q(w)$  позволяет заметно повысить надежность чамферного сопоставления, поскольку при неправильном пространственном преобразовании случайно совпадающие контурные точки, как правило, обладают разной ориентацией.

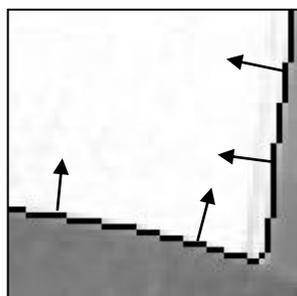


Рис. 2.25. Фрагмент изображения с выделенным контуром, для некоторых точек которого указано направление градиента

Однако использование информации об ориентации не делает контурные точки легко отождествляемыми и не избавляет от необходимости поиска в пространстве параметров преобразования.

#### *Построение структурных элементов путем сегментации контуров*

Контурные точки являются недостаточно уникальными. Если в качестве компонента взаимного пространственного преобразования изображений может выступать вращение, то даже с учетом приписанной контурной точке ориентации для этой точки сопряженной с ней может оказаться любая контурная точка второго изображения. В связи с этим необходимо построение описаний контуров, в которых были бы выделены некоторые характерные их точки или фрагменты. Подобные описания относятся к классу структурных описаний.

Широко используемыми являются описания контуров в виде аппроксимирующих их отрезков прямых линий и дуг эллипсов. Места соединения линий могут трактоваться как углы. В связи с этим задача построения структурных элементов на основе контуров может трактоваться как задача сегментации контуров, то есть задача разделения их на сегменты, каждый из которых описывается некоторой кривой. При этом контуры должны быть исходно представлены не как набор независимых точек, а как цепочки последовательно соединенных друг с другом точек (к примеру, на рис. 2.26 имеется девять цепочек).

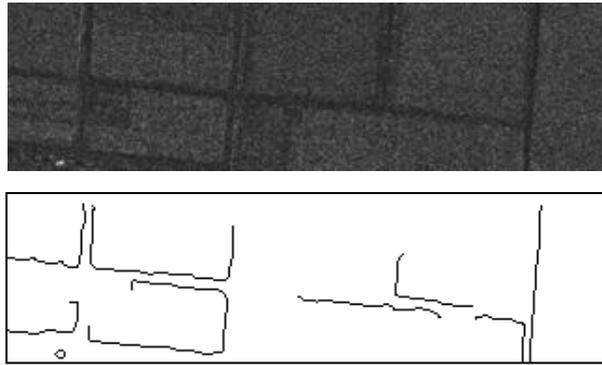


Рис. 2.26. Изображение и выделенные на нем контуры, состоящие из девяти цепочек

В качестве начального приближения можно рассматривать задачу независимого описания каждой из цепочек. В связи с этим мы будем полагать, что данный контур  $C = \{(x_i, y_i)\}_{i=1}^n$  представляет собой одну цепочку упорядоченных контурных точек (то есть точки  $(x_i, y_i)$  и  $(x_{i-1}, y_{i-1})$  являются четырех- или восьмисвязными для любого значения  $i$ ).

Рассмотрим сначала задачу аппроксимации прямой линией контура, состоящего из одного сегмента. Как правило, эта задача решается методом наименьших квадратов.

Пусть уравнение прямой задано в форме  $ax + by + c = 0$ , и необходимо определить значения параметров  $a$ ,  $b$  и  $c$ , при которых эта прямая наиболее точно аппроксимирует набор точек  $C = \{(x_i, y_i)\}_{i=1}^n$ . Расстояние от  $i$ -й точки до прямой определяется выражением

$$d_i = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}}.$$

Тогда требуется минимизировать критерий

$$D(a, b, c) = \sum_{i=1}^n \frac{(ax_i + by_i + c)^2}{a^2 + b^2}. \quad (2.24)$$

К сожалению, критерий (2.24) является нелинейным по параметрам  $a$  и  $b$ , поэтому поиск его минимума может быть затруднительным. Кроме того, если  $a^*x + b^*y + c^* = 0$  минимизирует данный критерий, то и  $(\alpha a^*)x + (\alpha b^*)y + (\alpha c^*) = 0$  при произвольном (отличном от нуля) значении  $\alpha$  также его минимизирует.

В связи с этим требуется наложить ограничения на параметры  $a$  и  $b$ , и решать задачу поиска условного экстремума. Удобным является ограничение  $a^2 + b^2 = 1$ , при выполнении которого необходимо минимизировать величину

$$D(a, b, c) = \frac{1}{n} \sum_{i=1}^n (ax_i + by_i + c)^2.$$

Решим эту задачу методом неопределенных множителей Лагранжа. Составим целевую функцию

$$L(a,b,c,\lambda) = \frac{1}{n} \sum_{i=1}^n (ax_i + by_i + c)^2 - \lambda(a^2 + b^2 - 1). \quad (2.25)$$

Необходимое условие экстремума выражается в виде системы уравнений:

$$\begin{cases} \frac{\partial L(a,b,c)}{\partial a} = \frac{2}{n} \sum_{i=1}^n (ax_i + by_i + c)x_i - 2\lambda a = 0 \\ \frac{\partial L(a,b,c)}{\partial b} = \frac{2}{n} \sum_{i=1}^n (ax_i + by_i + c)y_i - 2\lambda b = 0 \\ \frac{\partial L(a,b,c)}{\partial c} = \frac{2}{n} \sum_{i=1}^n (ax_i + by_i + c) = 0 \\ a^2 + b^2 = 1 \end{cases} \quad (2.26)$$

После элементарных преобразований система уравнений примет вид

$$\begin{cases} \left[ \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \lambda \right] a + \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) b + \left( \frac{1}{n} \sum_{i=1}^n x_i \right) c = 0 \\ \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) a + \left[ \left( \frac{1}{n} \sum_{i=1}^n y_i^2 \right) - \lambda \right] b + \left( \frac{1}{n} \sum_{i=1}^n y_i \right) c = 0 \\ \left( \frac{1}{n} \sum_{i=1}^n x_i \right) a + \left( \frac{1}{n} \sum_{i=1}^n y_i \right) b + c = 0 \\ a^2 + b^2 = 1 \end{cases}$$

Значения  $x_i$  и  $y_i$  известны. Переобозначим выражения  $M_{xx} = \frac{1}{n} \sum_{i=1}^n x_i^2$  и

т.д., тогда получим

$$\begin{cases} (M_{xx} - \lambda)a + M_{xy}b + M_x c = 0 \\ M_{xy}a + (M_{yy} - \lambda)b + M_y c = 0 \\ M_x a + M_y b + c = 0 \\ a^2 + b^2 = 1 \end{cases}$$

Исключая  $c$  из первых двух уравнений, получаем

$$\begin{cases} (M_{xx} - M_x^2 - \lambda)a + (M_{xy} - M_x M_y)b = 0 \\ (M_{xy} - M_x M_y)a + (M_{yy} - M_y^2 - \lambda)b = 0 \\ c = -(M_x a + M_y b) \\ a^2 + b^2 = 1 \end{cases}$$

Далее из первых двух уравнений несложно получить выражение  $[(M_{xy} - M_x M_y)(M_{xy} - M_x M_y) - (M_{yy} - M_y^2 - \lambda)(M_{xx} - M_x^2 - \lambda)]b = 0$ ,

Откуда

$$(M_{xy} - M_x M_y)(M_{xy} - M_x M_y) - (M_{yy} - M_y^2 - \lambda)(M_{xx} - M_x^2 - \lambda) = 0. \quad (2.27)$$

Решив данное квадратичное уравнение относительно  $\lambda$  и подставив полученные значения в систему, получим обычную систему линейных уравнений с нулевой правой частью и ограничением  $a^2 + b^2 = 1$ , дающим единственное решение для каждого из двух значений  $\lambda$  (одно из значений  $\lambda$  соответствует минимуму, а второе – максимуму критерия  $D(a, b, c)$ ). Таким образом, задача решена.

На рис. 2.27 представлен пример аппроксимации контуров, обнаруженных на изображении, посредством отрезков прямых линий с использованием описанного выше метода минимизации СКО.

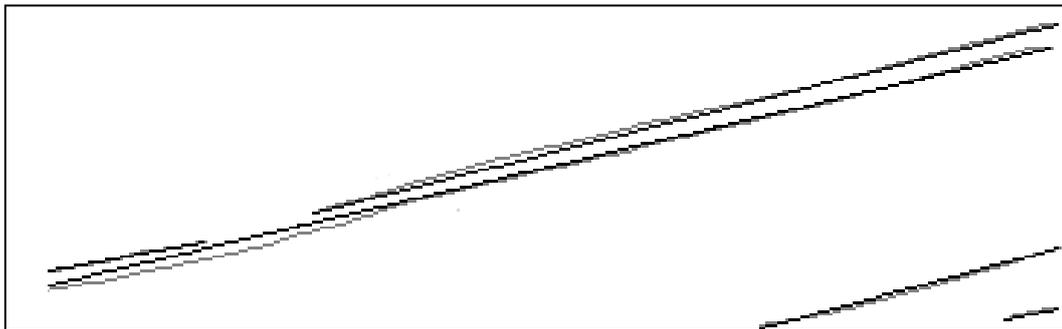


Рис. 2.27. Пример аппроксимации прямыми линиями некоторых контуров, выделенных на представленном рисунке, с помощью метода МНК

Рассмотрим теперь более сложный случай, при котором цепочка контура может содержать несколько сегментов, каждый из которых аппроксимируется собственной прямой. Основная трудность здесь заключается в определении границ сегментов и, соответственно, числа сегментов. То есть необходимо разбить контур  $\mathbf{C} = \{(x_i, y_i)\}_{i=1}^n$  на

совокупность сегментов  $\mathbf{C} = \bigcup_{k=1}^K \mathbf{S}_k, \mathbf{S}_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{n_k}$ , где  $(x_{k,i}, y_{k,i})$  –  $i$ -я

точка  $k$ -го сегмента  $\mathbf{S}_k$ ,  $n_k$  – число точек в  $k$ -м сегменте,  $\sum_{i=1}^K n_k = n$ .

Критерий СКО в данном случае не дает адекватного результата. Действительно, чем на большее число сегментов разделен контур, тем с меньшим значением СКО он может быть описан совокупностью линий.

Простейшим (но не лучшим) решением является введение порога на СКО, после превышения которого происходит разбиение текущего сегмента на два. Эту процедуру можно реализовать двумя путями.

1. Начиная с крайней точки контура, формировать сегмент, последовательно добавляя следующие точки, пока не будет превышен порог на СКО, после чего начать формировать следующий сегмент (при этом необходимо дополнительно уточнить точку, в которой начинается следующий сегмент, поскольку момент превышения порога на СКО будет с некоторым запаздыванием).
2. Взять весь контур в качестве начального сегмента и, если для него значение СКО превышает порог, разделить его на два. В качестве точки разделения выбрать точку, которая дает минимальную суммарную ошибку по двум формируемым сегментам.

Данные процедуры в определенной степени работают, однако введение порога на СКО является недостаточно гибким решением.

Другой подход заключается в том, чтобы обнаруживать сами границы сегментов. В этих точках происходит наиболее быстрое изменение направления контура, в связи с чем их можно искать как точки, в которых значение кривизны контура имеет локальный максимум. Существуют различные способы оценки кривизны контура или ее аналогов. Следующий простейший способ проиллюстрирован на рис. 2.28.

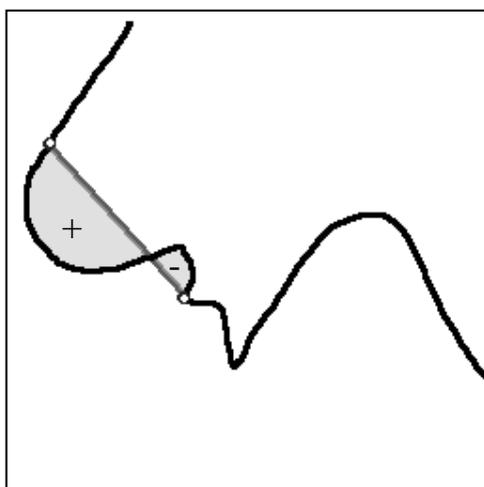


Рис. 2.28. Иллюстрация к вычислению параметра локальной кривизны контура

Для вычисления параметра, определяющего локальную кривизну контура в текущей точке, выбирается точка контура, расположенная по контуру на некотором фиксированном удалении от текущей (этим определяется окно усреднения). Эти две точки соединяются линией. Вычисляются площади, областей, ограниченных контуром по обе стороны от этой линии. При этом площади по одну сторону от линии прибавляются к общей площади, а по другую – вычитаются из нее (благодаря этому для дуги

окружности будет получено большое значение площади, а для контура, из-за шумов колеблющегося вокруг прямой линии, – малое). Затем полученное число делится на длину соединяющей линии.

Подход на основе точек максимальной кривизны тоже не дает универсального решения проблемы определения границ сегментов, поскольку не различает локальных всплесков кривизны, вызванных шумами, от их регулярных изменений. Для устранения шумов поиск локальных максимумов кривизны может производиться в большем окне, однако это может привести к потере деталей на контурах малоразмерных объектов.

На рис. 2.29 приведена иллюстрация к проблеме выделения сегментов на контуре.

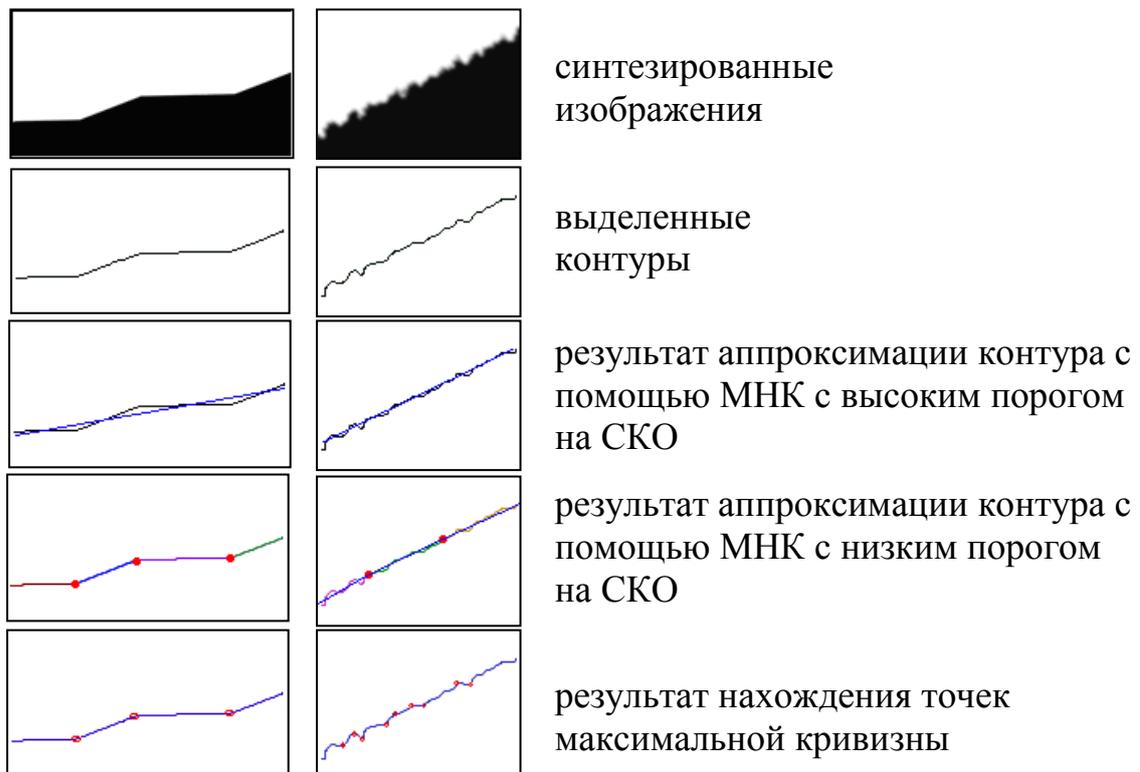


Рис. 2.29. Иллюстрация проблемы определения числа сегментов при решении задачи сегментации контуров с использованием критерия СКО с ограничением по порогу и точек максимальной кривизны

Проблема определения числа сегментов, на которые следует разбить контур, в определенной степени решается с помощью принципа минимальной длины описания (МДО). Согласно этому принципу следует выбрать ту модель данных, при которой сумма длины модели и длины данных, описанных с помощью модели, является минимальной. В данном случае минимизируется длина описания невязок  $d_i$  и параметры линий  $a_k, b_k$  и  $c_k$ , описывающих каждый сегмент

$$L(\{a_k, b_k, c_k\}_{i=1}^K) = n_k \log_2 D_k(a_k, b_k, c_k) + \sum_{k=1}^K \frac{m_k}{2} \log_2 n_k + \log_2 K, \quad (2.28)$$

где  $D_k(a_k, b_k, c_k)$  – дисперсия невязок точек  $k$ -го сегмента при данном значении параметров,  $m_k = 2$  – число независимых параметров линии, описывающей  $k$ -й сегмент (если описание сегмента ведется не только отрезками прямых линий, но и другими структурными элементами, то величина  $m_k$  будет разной для разных сегментов).

Таким образом, этот критерий учитывает не только величины невязок, с которыми структурные элементы аппроксимируют контур, но и число этих элементов. При этом если добавление нового сегмента приводит лишь к незначительному понижению дисперсии невязок, то в соответствии с этим критерием, новый сегмент вводить нерационально. Это позволяет различить два случая, представленных на рис. 2.29: случай изломанного контура, состоящего из нескольких выраженных сегментов, и случай прямого, но зашумленного контура. При этом может использоваться алгоритм последовательного расщепления контура на новые сегменты, либо последовательного слияния сегментов, вначале представляющих собой элементарные прямые отрезки.

Таким образом, проблему сегментации контуров с формированием отрезков прямых линий в качестве структурных элементов, можно решать с помощью критерия СКО с ограничением по порогу, путем поиска точек максимальной кривизны, а также с помощью критерия МДО. Сходным образом в дополнение к отрезкам прямых линий можно осуществлять использовать сегменты кривых второго порядка (как показывает опыт разработчиков систем компьютерного зрения, использование кривых более высоких порядков, за исключением специфических случаев, оказывается неэффективным).

Процедуры построения структурных элементов на основе контуров не ограничиваются решением задачи сегментации контуров. Кроме этого часто оказывается необходимым заполнять разрывы контуров, возникающие из-за низкого контраста или шумов (см. рис. 2.30).

Процедуры заполнения разрывов, как правило, имеют много общего с процедурой сегментации, в которой производится последовательное объединение сегментов. Здесь, однако, объединяются сегменты, принадлежащие не одной, а разным контурным цепочкам. При этом условием, при выполнении которого производится проверка возможности объединения сегментов, является близость концов сегментов, расположенных на разных контурных цепочках.

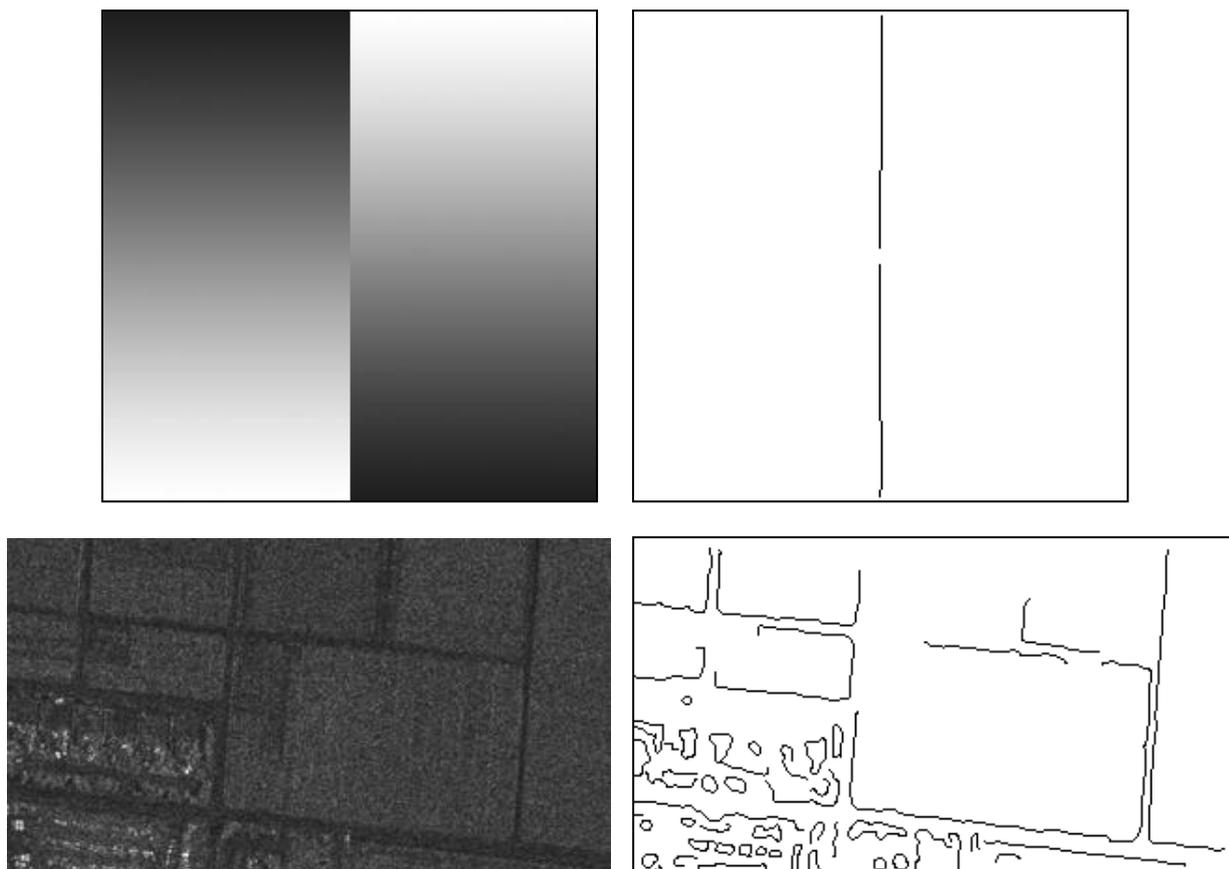


Рис. 2.30. Примеры разрывов контуров на синтезированном изображении в области с низким значением модуля градиента и на реальном радиолокационном изображении (в правой верхней части) вследствие спекл-шума

*Построение геометрических элементов на основе преобразования Хафа*

Пусть задано уравнение  $\varphi(\mathbf{x}, \mathbf{w}) = 0$ , связывающее переменные, из которых составлен вектор  $\mathbf{x}$ , и зависящее от вектора параметров  $\mathbf{w}$ . При этом дан набор значений  $\{\mathbf{x}_i\}_{i=1}^n$ , часть которых удовлетворяет уравнению  $\varphi(\mathbf{x}, \mathbf{w}) = 0$  при одном и том же значении вектора параметров  $\mathbf{w}^*$ , однако это значение  $\mathbf{w}^*$  неизвестно. Остальные же значения  $\mathbf{x}_i$  либо удовлетворяют уравнению  $\varphi(\mathbf{x}, \mathbf{w}) = 0$  при других значениях  $\mathbf{w}$ , либо не удовлетворяют ему вообще. При этом неизвестно не только значение  $\mathbf{w}^*$ , но и какое именно подмножество множества  $\{\mathbf{x}_i\}_{i=1}^n$  удовлетворяет уравнению при искомом значении  $\mathbf{w}^*$ . Таким образом, требуется найти как соответствующее подмножество, так и значение  $\mathbf{w}^*$ .

Задача поиска на изображении прямых линий или дуг окружностей и эллипсов может быть поставлена в такой форме в случае, когда произведено не выделение связанных контуров, а обнаружение несвязанных краевых точек в рамках признакового подхода. Поскольку краевые точки не сгруппированы, неизвестно, какие из них относятся к тому или иному отрезку прямой или

эллипса. Все точки заданы в одном массиве  $\mathbf{C} = \{(x_i, y_i)\}_{i=1}^n$ , поэтому для построения соответствующих структурных элементов (в данном случае называемых геометрическими элементами, поскольку они не образуют какой-либо структуры) необходимо не только определить параметры соответствующего элемента, но и выделить подходящие краевые точки.

Данная проблема решается с помощью т.н. *преобразования Хафа*. Для пояснения сущности этого преобразования вернемся к уравнению  $\varphi(\mathbf{x}, \mathbf{w}) = 0$ . Одно значение вектора  $\mathbf{x}$  не позволяет однозначно определить значение  $\mathbf{w}^*$ , при котором будет выполнено это уравнение, поскольку задает лишь одно ограничение. Пусть размерность вектора  $\mathbf{w}$  равна  $M$ . Тогда нам потребовалось бы не менее  $M$  векторов из множества  $\{\mathbf{x}_i\}_{i=1}^n$ , чтобы установить нужное значение  $\mathbf{w}^*$ . Однако нам не известно, какие векторы из этого множества соответствуют искомому значению  $\mathbf{w}^*$ , поэтому нам пришлось бы перебирать все сочетания, состоящие из  $M$  векторов для получения гипотез о возможных значениях  $\mathbf{w}^*$ , и каким-то образом выбирать из всех гипотез лучшую. Для этого потребовалось бы более чем  $C_n^M$  операций, что может быть весьма большим числом.

При преобразовании Хафа рассматриваются не полные сочетания по  $M$  векторов, а сокращенные сочетания, включающие меньшее число векторов. В частности, могут рассматриваться лишь отдельные элементы  $\mathbf{x}_i$ . Каждый такой элемент при подстановке в уравнение  $\varphi(\mathbf{x}_i, \mathbf{w}) = 0$  задает некоторую поверхность (размерности  $M-1$ ) в пространстве параметров  $\mathbf{w}$ . Аналогично, пара элементов  $\mathbf{x}_i$  и  $\mathbf{x}_j$  задает в пространстве параметров многообразие  $M-2$  размерности. При этом все поверхности в пространстве параметров, задаваемые векторами из искомого подмножества множества  $\{\mathbf{x}_i\}_{i=1}^n$  должны пересекаться в одной и той же точке (при отсутствии погрешностей), соответствующем искомому значению  $\mathbf{w}^*$ .

Для поиска точки пересечения используется метод «голосования», состоящий из следующих шагов:

1. Разбить пространство параметров на ячейки, в каждой из которых хранится число голосов  $\nu(\mathbf{w})$ , где  $\mathbf{w}$  – значение параметров, соответствующее данной ячейке. В начале работы метода  $(\forall \mathbf{w}) \nu(\mathbf{w}) = 0$ .
2. Для каждой точки  $\mathbf{x}_i$  определить множество ячеек, для которых выполняется уравнение  $\varphi(\mathbf{x}_i, \mathbf{w}) = 0$  и увеличить для них значение  $\nu(\mathbf{w})$ :  $(\forall \mathbf{w}) \varphi(\mathbf{x}_i, \mathbf{w}) = 0 \Rightarrow \text{inc}(\nu(\mathbf{w}))$ .
3. Найти максимум в массиве ячеек  $\nu(\mathbf{w})$  и определить соответствующее значение  $\mathbf{w}^*$ , по которому установить подмножество  $\{\mathbf{x}_i^*\}_{i=1}^n$  тех векторов из множества  $\{\mathbf{x}_i\}_{i=1}^n$ , которые «голосовали» за эту ячейку.

4. Повторить процедуру для множества  $\{\mathbf{x}_i\}_{i=1}^n \setminus \{\mathbf{x}_i^*\}_{i=1}^{n^*}$  с исключенными «проголосовавшими» векторами.

В пункте 2 можно перебирать пары  $(\mathbf{x}_i, \mathbf{x}_j)$  и увеличивать значение  $v(\mathbf{w})$  по условию  $\varphi(\mathbf{x}_i, \mathbf{w}) = 0 \ \& \ \varphi(\mathbf{x}_j, \mathbf{w}) = 0$ . Это позволяет снизить размерность соответствующего многообразия в пространстве параметров, однако увеличивает количество «голосующих» объектов. К примеру, при построении линий одна точка  $(x_i, y_i)$  в пространстве параметров задает кривую  $x_i \cos \varphi + y_i \sin \varphi + c = 0$ , где  $\varphi$  – азимут линии, в то время как пара точек  $(x_i, y_i)$  и  $(x_j, y_j)$  однозначно задает значение  $\varphi$  и  $c$ , то есть «голосует» только за одну ячейку.

Могут перебираться и тройки векторов, но применяется это редко. При этом при использовании пар и троек векторов часто выполняется не полный перебор их возможных комбинаций, а производится случайный выбор фиксированного числа комбинаций, что повышает скорость, но уменьшает надежность работы метода.

Здесь также может использоваться подход с переменным разрешением. В рамках этого подхода пространство параметров сначала разбивается на крупные ячейки, среди которых выбирается ячейка с наибольшим значением  $v(\mathbf{w})$ . После этого рассматривается лишь фрагмент пространства параметров, соответствующий этой ячейке, которая разбивается на более мелкие ячейки. Таким образом, решение постепенно уточняется.

На рис. 2.31 приведен пример обнаружения прямых линий (изображенных утолщенными серыми линиями) на основе контурных точек с использованием преобразования Хафа. Как видно из рисунка, выделяются наиболее длинные прямые линии. Поскольку в данном подходе не используется информация о принадлежности краевых точек одним и тем же цепочкам контуров, метод не чувствителен к разрывам контуров, но при этом в качестве результатов его работы могут строиться линии, включающие случайные лишние точки.

Преобразование Хафа хорошо применимо при поиске четко выраженных линий или окружностей, но не для построения структурных описаний изображений, поскольку часто структурные элементы соответствуют достаточно малым сегментам контуров, которые не дают достаточного количества «голосов» при преобразовании Хафа для их обнаружения.

С помощью преобразования Хафа можно также искать и более сложные геометрические элементы – квадраты и прямоугольники. Однако увеличение числа параметров в искомой модели существенным образом сказывается на скорости работы метода и требуемой памяти. Простейшая реализация метода требует порядка  $N^M$  байт, где  $N$  – размер изображения, а  $M$  – размерность вектора параметров; уже при  $M=4$  требования к памяти становятся чрезвычайно жесткими, и оказывается необходимым применять подход с переменным разрешением.

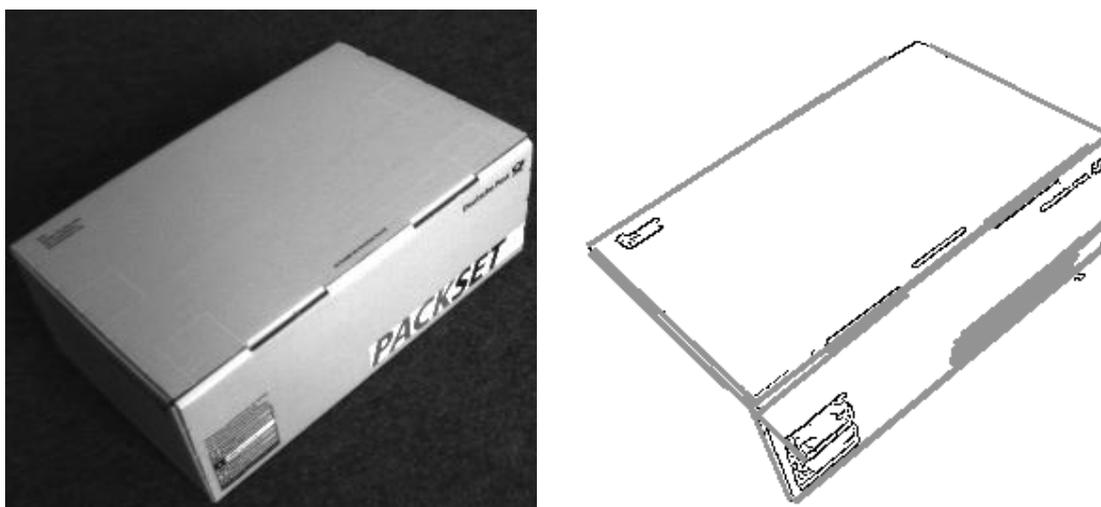


Рис. 2.31. Пример обнаружения прямых линий на основе контуров (изображенных тонкими черными линиями), выделенных на изображении

*Структурное сопоставление изображений с применением преобразования Хафа*

Рассмотрим проблему структурного сопоставления изображений. Пусть даны два набора структурных элементов  $\{s_{1,k}\}_{k=1}^{K_1}$  и  $\{s_{2,k}\}_{k=1}^{K_2}$ . В качестве структурных элементов могут выступать отрезки прямых линий, дуги окружностей, точки максимальной кривизны и т.д. Несмотря на различие типов элементов, информацию о структурных элементах можно единообразно представить в виде векторов некоторых значений, таких как координаты центра элемента, азимут ориентации элемента (направление линии или направление биссектрисы угла), размеры элемента (длина линии, раствор угла) и т.д. Пространственное совмещение элементов выполняется на основе их координат, а прочие параметры элементов могут использоваться для проверки возможности их сопоставления. Рассмотрим сначала случай, при котором элементы заданы только своими координатами, тогда  $s_{i,k} = (x_{i,k}, y_{i,k})$ .

Заранее неизвестно, какие элементы первого изображения сопряжены с какими элементами второго изображения. Координаты пары сопряженных элементов  $(x_{1,i}, y_{1,i})$  и  $(x_{2,j}, y_{2,j})$  должны удовлетворять двум уравнениям  $\varphi_x(x_{1,i}, y_{1,i}, \mathbf{w}) - x_{2,j} = 0$  и  $\varphi_y(x_{1,i}, y_{1,i}, \mathbf{w}) - y_{2,j} = 0$ . Одной пары сопряженных точек, как правило, недостаточно для однозначного восстановления параметров пространственного преобразования: если размерность вектора параметров равна  $M$ , то необходимо не менее  $M/2$  сопоставлений. Как и при построении структурных элементов, можно перебирать все наборы из  $M/2$  пар точек  $((x_{1,i}, y_{1,i}), (x_{2,j}, y_{2,j}))$ , для каждого набора определять параметры пространственного преобразования и устанавливать некоторым образом качество этого набора.

При структурном сопоставлении такой перебор более приемлем, чем при построении структурных элементов, поскольку число структурных элементов значительно меньше числа контурных точек, и этот способ сопоставления мы также рассмотрим. Однако здесь может быть применено и преобразование Хафа. Действительно, каждая пара  $((x_{1,i}, y_{1,i}), (x_{2,j}, y_{2,j}))$  задает многообразие размерности  $M-2$  в пространстве признаков. Таким образом, «голосует» за некоторое число ячеек в пространстве параметров, и ячейка с максимальным значением  $v(\mathbf{w})$  соответствует искомому значению параметров пространственного преобразования  $\mathbf{w}^*$ , а пары элементов, «голосовавших» за эту ячейку, определяют пару сопряженных точек.

К сожалению, как и при построении структурных элементов, в случае сопоставления изображений применение преобразования Хафа наталкивается на определенные трудности при больших значениях  $M$  (больше трех или четырех). Тем не менее, для таких пространственных преобразований, как преобразования сдвига или движения, преобразование Хафа вполне может быть применено для решения задачи сопоставления.

Так, например, на рис. 2.32 представлен пример преобразования Хафа, выполненного по совокупности центров структурных элементов (или точек интереса). Изображения связаны только преобразованием сдвига, поэтому пространство параметров двумерно, и не представляет трудности заполнить массив  $v(\mathbf{w})$ : любые два сопоставляемых элемента определяют одну точку в пространстве параметров.

При использовании преобразования Хафа следует иметь в виду возможные погрешности в определении координат структурных элементов. Это приводит к тому, что сопряженные точки задают поверхности в пространстве параметров, которые не пересекаются в точке  $\mathbf{w}^*$ , а лишь проходят вблизи нее.

При выборе слишком малого размера ячеек может оказаться, что все значения  $v(\mathbf{w})$  малы, и максимум не может быть локализован (этот эффект будет присутствовать и при построении структурных элементов с помощью преобразования Хафа). В связи с этим необходимо либо использовать ячейки большего размера (что может, однако, привести к тому, что в одну ячейку попадет несколько локальных максимумов), либо осуществлять «голосование» не только за ячейки, через которые проходит соответствующая поверхность, но и за соседние с ними ячейки (возможно, с меньшим весом).

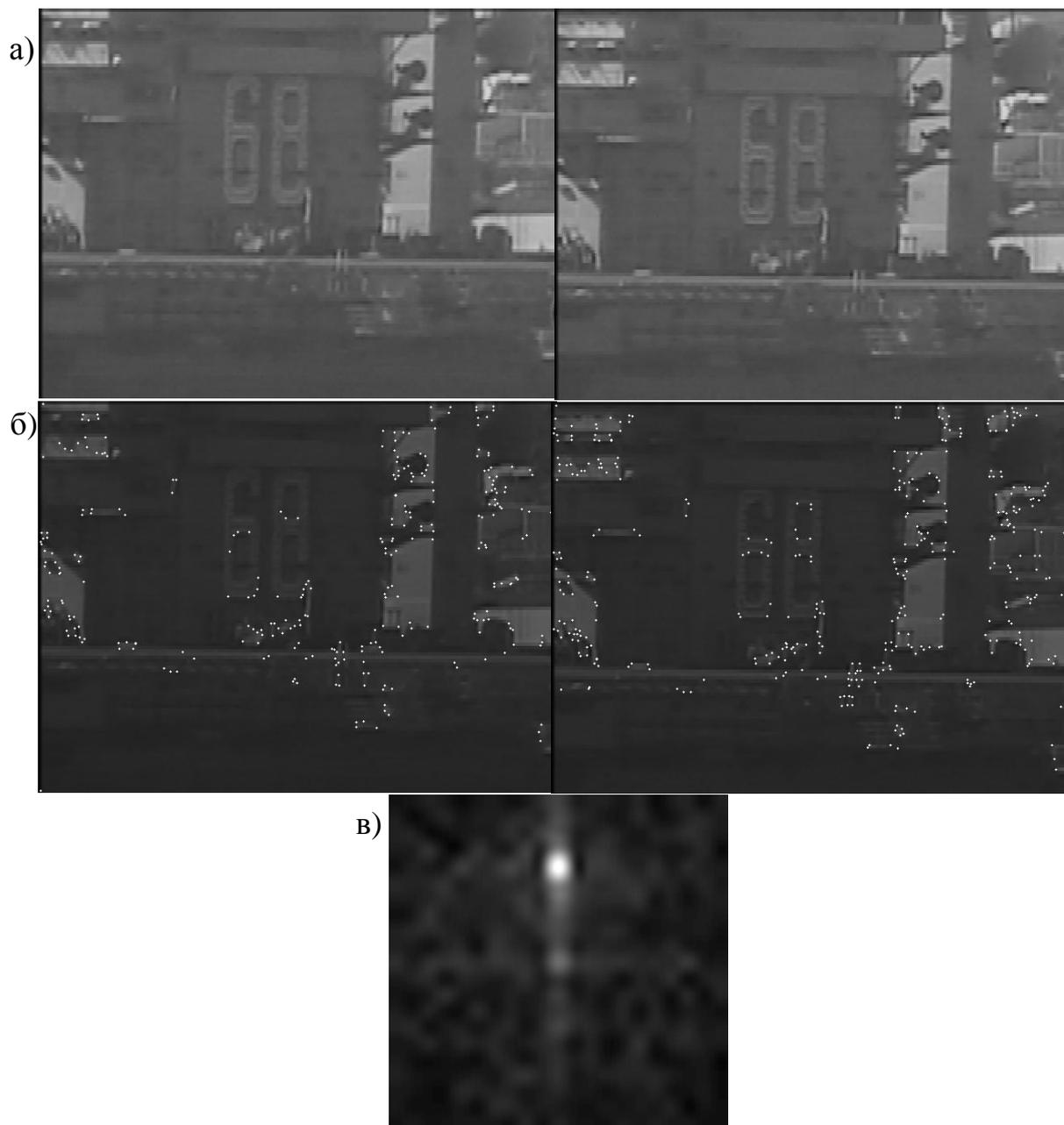


Рис. 2.32. а) Два изображения, снятые камерой, помещенной на нестабилизированную платформу, и различающиеся сдвигом; б) изображения с нанесенными на них точками интереса; в) результат преобразования  $X_0$  для точек интереса; максимум соответствует истинному вектору смещения  $(0, 20)$

### *Поиск по дереву вариантов*

Наиболее мощным и гибким методом структурного сопоставления изображений является перебор комбинаций возможных соответствий структурных элементов. К примеру, на рис. 2.33 представлены два изображения с выделенными на них структурными элементами. Рассмотрим пронумерованные углы, присутствующие на этих изображениях. Существует множество комбинаций возможных соответствий, среди которых есть истинная комбинация  $((1, 7), (3, 1), (4, 2), \dots)$ , а также прочие комбинации, причем, как удовлетворяющие единому пространственному преобразованию, например,  $((2, 7), (8, 1), (7, 2), (6, 3))$ , так и не удовлетворяющие такому

преобразованию, например,  $((1, 3), (2, 7), (3, 4), (7, 5))$ . Комбинации могут также содержать и разное число сопоставлений. Всего возможно  $K_1 K_2$  различных сопоставлений, следовательно, если не учитывать никаких ограничений, возможно  $(K_1 K_2)!$  их комбинаций. Это число крайне велико, однако все комбинации перебирать не имеет смысла.

Как правило, перебор осуществляется в форме поиска по дереву вариантов соответствий. Идея здесь заключается в следующем. Первый элемент первого изображения может быть поставлен в соответствие любому элементу второго изображения. То есть у нас есть возможность пойти по одной из  $K_2$  ветвей, выходящих из корня дерева, в котором никакие соответствия не определены, и ведущих в  $K_2$  узлов следующего уровня.

В каждом узле следующего уровня второй элемент первого изображения может быть сопоставлен  $K_2 - 1$  узлам второго изображения (то есть будет порядка  $(K_2 - 1)K_2$  узлов второго уровня). И так далее, пока не будет достигнут один из листьев дерева, в котором все соответствия установлены. В действительности, необходимо для каждого узла добавить еще одну выходящую ветвь, которая определяет отсутствие соответствия текущего элемента первого изображения какому-либо элементу второго изображения. На рис. 2.34 представлен фрагмент дерева варианта соответствий структурных элементов, приведенных на рис. 2.33.

Как видно, задача структурного сопоставления сводится к одной из задач эвристического программирования, связанной с поиском на дереве вариантов. Для полной постановки задачи необходимо указать критерий качества решения, который может быть вычислен в листьях. Для этого рассмотрим некоторую комбинацию соответствий  $\{(x_{1,i}, y_{1,i}), (x_{2,j}, y_{2,j})\}_{i=1}^d$ . По этой комбинации можно с помощью метода МНК определить оптимальные параметры пространственного преобразования путем минимизации следующей целевой функции:

$$D^2(\mathbf{w}) = \frac{1}{d} \sum_{i=1}^d \left[ (\varphi_x(x_{1,i}, y_{1,i}, \mathbf{w}) - x_{2,i})^2 + (\varphi_y(x_{1,i}, y_{1,i}, \mathbf{w}) - y_{2,i})^2 \right]. \quad (2.29)$$

Сам по себе критерий  $D(\mathbf{w})$  не может служить качеством комбинации соответствий, поскольку, чем меньше размер комбинации  $d$ , тем, как правило, меньше значение  $D(\mathbf{w})$ . В то же время, нас интересуют комбинации как можно большего размера. В связи с этим, критерий качества должен включать как  $d$ , так и  $D(\mathbf{w})$ . В простейшем случае эвристически вводятся критерии вида

$$Q(\mathbf{w}) = d - D(\mathbf{w})/a, \quad Q'(\mathbf{w}) = d - D^2(\mathbf{w})/a^2 \quad \text{или} \quad Q''(\mathbf{w}) = d(1 - D(\mathbf{w})/a),$$

где настроенный коэффициент  $a$  характеризует вес, с которым учитывается среднеквадратичная ошибка отображения сопряженных точек по сравнению с размером комбинации. Наиболее приемлемый вид критерия существенно зависит от способа построения структурных элементов и содержания изображений.

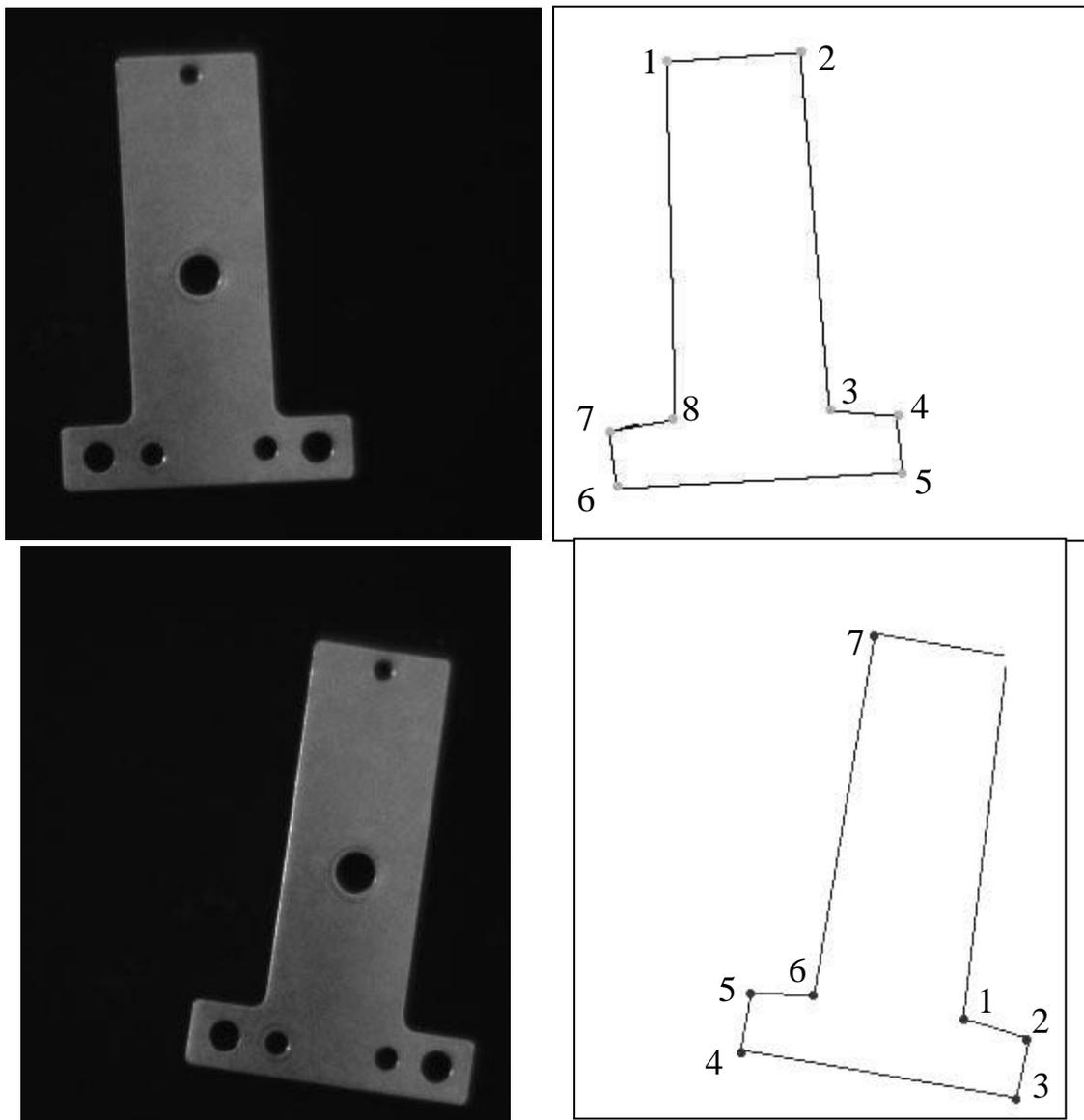


Рис. 2.33. Пример изображений деталей на ленте конвейера с выделенными в качестве структурных элементов углами, которым присвоены индексы

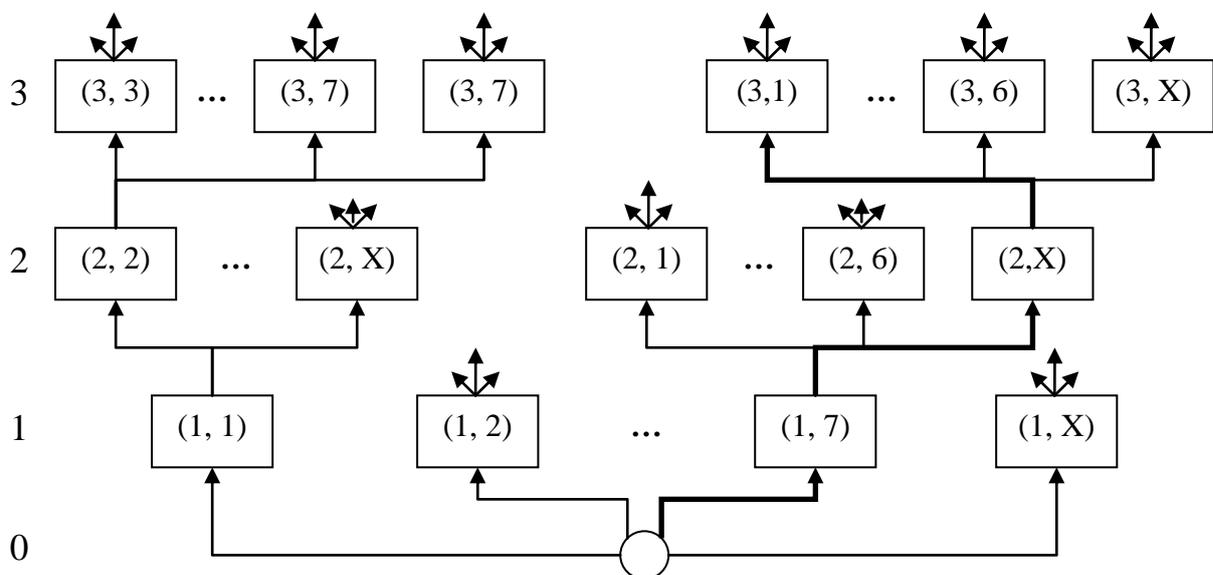


Рис. 2.34. Фрагмент дерева вариантов; утолщенной линией показан путь на дереве, соответствующий правильному сопоставлению элементов

При решении задачи поиска на дереве вариантов в эвристическом программировании в качестве центральной проблемы рассматривается проблема комбинаторного взрыва, которая, очевидно, имеет место и здесь. Для преодоления этой проблемы предлагается использовать эвристики, позволяющие отсекалть неперспективные ветви дерева.

К сожалению, классические способы сокращения перебора, разработанные в эвристическом программировании, такие как методы  $N$ -наилучшего направленного сокращения или альфа-отсечения здесь оказываются недостаточно эффективными. Это связано с тем, что в задаче сопоставления изображений оказывается затруднительным ввести оценивающую функцию для промежуточных узлов дерева, которая бы монотонно сходилась к значению критерия качества в листьях (хотя этот вопрос на настоящий момент исследован плохо). Вместо этого при решении задачи структурного сопоставления изображений используются специфические эвристики поиска.

1. Основным способом ограничения перебора является предварительная оценка пространственного преобразования. Пусть размерность вектора параметров  $\mathbf{w}$  равно  $M$ . Тогда достаточно  $M/2$  выбранных соответствий для однозначной оценки вектора параметров  $\mathbf{w}$ . Таким образом, эта оценка может быть уже произведена на уровне  $M/2$  дерева вариантов. После этого последующие ветки могут эффективно отсекалться. Например, для преобразования подобия  $M=4$ , то есть его параметры могут быть оценены для каждого узла второго уровня (не считая узлов, путь к которым проходит через узлы с неустановленными соответствиями). К примеру, на рис. 2.34, самая левая из изображенных ветвей на уровне 2 определяет соответствия (1, 1) и (2, 2), которые задают такое пространственное преобразование (верх левого изображения совмещается с низом правого изображения), что ни для каких других структурных элементов соответствия определены быть не могут, и соответствующие ветки должны быть отсечены. Следует, однако, иметь в виду, что по  $M/2$  соответствий параметры пространственного преобразования могут определяться с большой ошибкой, поэтому критерий отсечения последующих ветвей на основе предварительно оцененного преобразования не должен быть слишком жестким.
2. Некоторые возможные соответствия элементов первого и второго изображения могут исключаться на основе дополнительных признаков. К примеру, можно потребовать, чтобы растворы сопоставляемых углов или длины сопоставляемых линий отличались не слишком сильно. Кроме того, любые пары сопоставленных элементов на одном и втором изображении должны иметь мало отличающиеся относительные азимуты ориентации (см. рис. 2.35). Возможность введения подобных ограничений, однако, существенно зависит от того, насколько структурные элементы устойчиво выделяются.

Оценим сложность алгоритма поиска по дереву вариантов при использовании предварительной оценки пространственного преобразования. Дерево можно разделить на две части: в первой части не осуществляется отсечение ветвей (мы рассматриваем простейший случай, когда элементы не имеют дополнительных признаков), а во второй части осуществляется отсечение ветвей на основе оцененного пространственного преобразования.

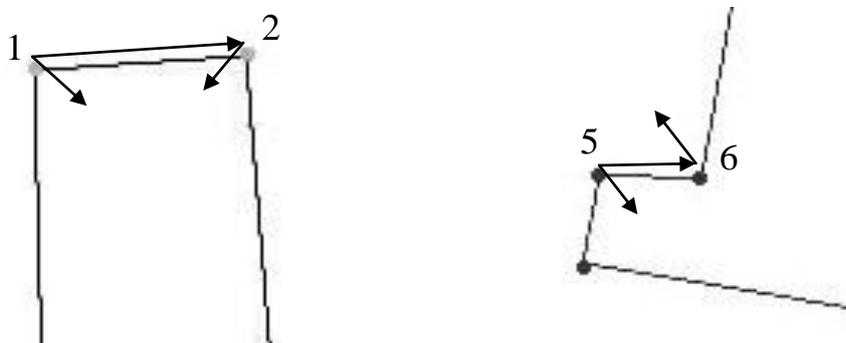


Рис. 2.35. Фрагменты структурных описаний, представленных на рис. 2.33: угол между вектором ориентации элемента 1 и вектором направления на элемент 2 совпадает с углом между вектором ориентации элемента 5 и вектором направления на элемент 6; в то же время этот же угол для элемента и элемента 6 разный (одинаковый по величине, но разный по знаку), что позволяет заключить о невозможности одновременного сопоставления обоих элементов пар

Размер начала дерева  $K_2(K_2 - 1) \dots (K_2 - M/2) \approx K_2^{M/2}$ . Далее для каждого из  $K_1 - M/2$  элементов первого изображения необходимо (по оцененному по начальной части дерева пространственному преобразованию) осуществить выбор лучшего элемента второго изображения. Для этого необходимо порядка  $K_1 K_2$  операций, выполняемых для каждого конечного узла начальной части дерева. В итоге потребуется порядка  $K_1 K_2^{1+M/2}$  операций. В действительности, число операций может оказаться несколько большим, если допустить возможность того, что элементы первого изображения могут не соответствовать никаким элементам второго изображения.

Из-за большого числа структурных элементов и высокой сложности перебора, как правило, приходится выбирать ограниченного число элементов, участвующих в структурном сопоставлении. При этом выбираются элементы, выделяющиеся наиболее надежно и обладающие наиболее выраженными характеристиками. К примеру, если речь идет об отрезках прямых линий, построенных на основе контуров, то могут выбираться линии, разделяющие области, наиболее сильно отличающиеся по яркости, а также наиболее длинные линии и линии, описывающие контур с наименьшим значением СКО. Кроме того, наиболее выделяющиеся элементы могут помещаться в начало дерева вариантов, тогда начальная часть дерева вариантов, в которой происходит формирование предварительной оценки параметров пространственного преобразования, будет иметь меньшие размеры.

Сокращение перебора может достигаться путем повышения уникальности структурных элементов, что может осуществляться несколькими путями. Во-первых, элементы могут снабжаться большим числом дополнительных признаков, на основе которых можно было бы попарно сопоставлять элементы и исключать неперспективные соответствия. Здесь основная трудность заключается в том, чтобы добиться инвариантности этих признаков к геометрическим и яркостным преобразованиям изображений.

Во-вторых, на основе производных структурных элементов могут формироваться составные структурные элементы, которые будут различаться по своим типам (например, треугольник не может быть поставлен в соответствие параллелограмму, а Т-образное соединение – U-образному соединению). В предельном случае на изображениях может производиться распознавание конкретных объектов, которые, выступая при сопоставлении изображений вместо структурных элементов, могут быть сопоставлены с гораздо меньшей неоднозначностью. Подобные методы сопоставления, основанные на знаниях, хотя и существуют, но каждый из них имеет узкую сферу применения, поскольку на настоящий момент не существует универсальных систем распознавания изображений.

Помимо методов решения проблемы структурного сопоставления изображений с помощью поиска по дереву вариантов или преобразования Хафа существуют и другие методы, например, основанные на нейронных сетях, генетических алгоритмах и т.д. Однако эти методы отличаются лишь механизмом поиска, а не существом выполняемого структурного сопоставления. Весьма оригинальным, но сейчас мало используемым подходом к структурному сопоставлению является подход на основе формальных грамматик. Их рассмотрение, однако, выходит за рамки курса.

### **Вопросы и упражнения**

1. Каким образом определяется преобразование расстояния?
2. В чем заключается метод чамферного сопоставления?
3. Как может применяться информация об ориентации граничных точек при чамферном сопоставлении?
4. Какие методы построения структурных элементов вы знаете?
5. В чем заключается проблема сегментации контуров с использованием критерия СКО? Как эта проблема может решаться?
6. В чем заключается идея преобразования Хафа?
7. Как преобразование Хафа применяется в задачах сопоставления изображений? Каковы его ограничения?
8. Как строится дерево вариантов соответствий в задаче структурного сопоставления?
9. Какие способы отсечения ветвей на дереве вариантов?
10. Какова сложность перебора в методе структурного сопоставления в различных случаях?

### 3. Восстановление трехмерной структуры сцены

#### 3.1. Проблема стереозрения

Ранее были рассмотрены методы сопоставления изображений, преимущественно нацеленные на восстановление глобального взаимного преобразования, характерного для плоских сцен. Если же с двух ракурсов снята существенно трехмерная сцена, то взаимное преобразование изображений не будет описываться единой глобальной моделью, поскольку изменение координат каждой точки будет зависеть от расстояния от нее до камеры. Здесь мы рассмотрим проблему сопоставления трехмерных сцен.

В результате такого сопоставления может быть восстановлена информация о дальности до точек сцены (как правило, в этом и заключается цель сопоставления подобных изображений). Существует множество других методов восстановления трехмерной информации, не опирающихся на геометрическое преобразование изображений при смене ракурса съемки, однако их краткий обзор будет выполнен позднее.

В простейшем случае при сопоставлении двух изображений некоторой трехмерной сцены, снятой с разных ракурсов, взаимное положение камер при получении сопоставляемых изображений известно. В важном частном случае положение камер отличается лишь сдвигом в направлении, перпендикулярном оптической оси, которая для обеих камер имеет одинаковое направление. Задача сопоставления изображений в этом случае называется задачей *стереозрения*, а расстояние между камерами – *стереобазой*.

Стереозрение является эффективным средством восстановления трехмерной формы объектов и расстояний до них, что широко используется в биологических системах зрительного восприятия. В частности, осуществляющееся в зрительной коре объединение информации, поступающей от двух глаз, помогает человеку воспринимать мир трехмерным (хотя это и не единственный работающий механизм).

Идея стереозрения в технических системах может применяться по отношению к самым разным объектам. К примеру, стереопары могут содержать изображения микроскопических объектов, сцен внутри помещений, аэрокосмические снимки и т.д. На рис. 3.1 приведены примеры некоторых стереопар с различным содержанием. Несмотря на существенное различие масштабов, задачи стереозрения для всех представленных случаев является во многом сходными.

Посмотрим, как связано различие в положении объекта на изображениях стереопары с расстоянием до него. Как видно из рис. 3.2, угол, под которым виден некоторый объект, зависит как от расстояния до него (вернее, положения вдоль оси  $Z$ ), так и от его положения вдоль оси  $X$ , перпендикулярной оптической оси.

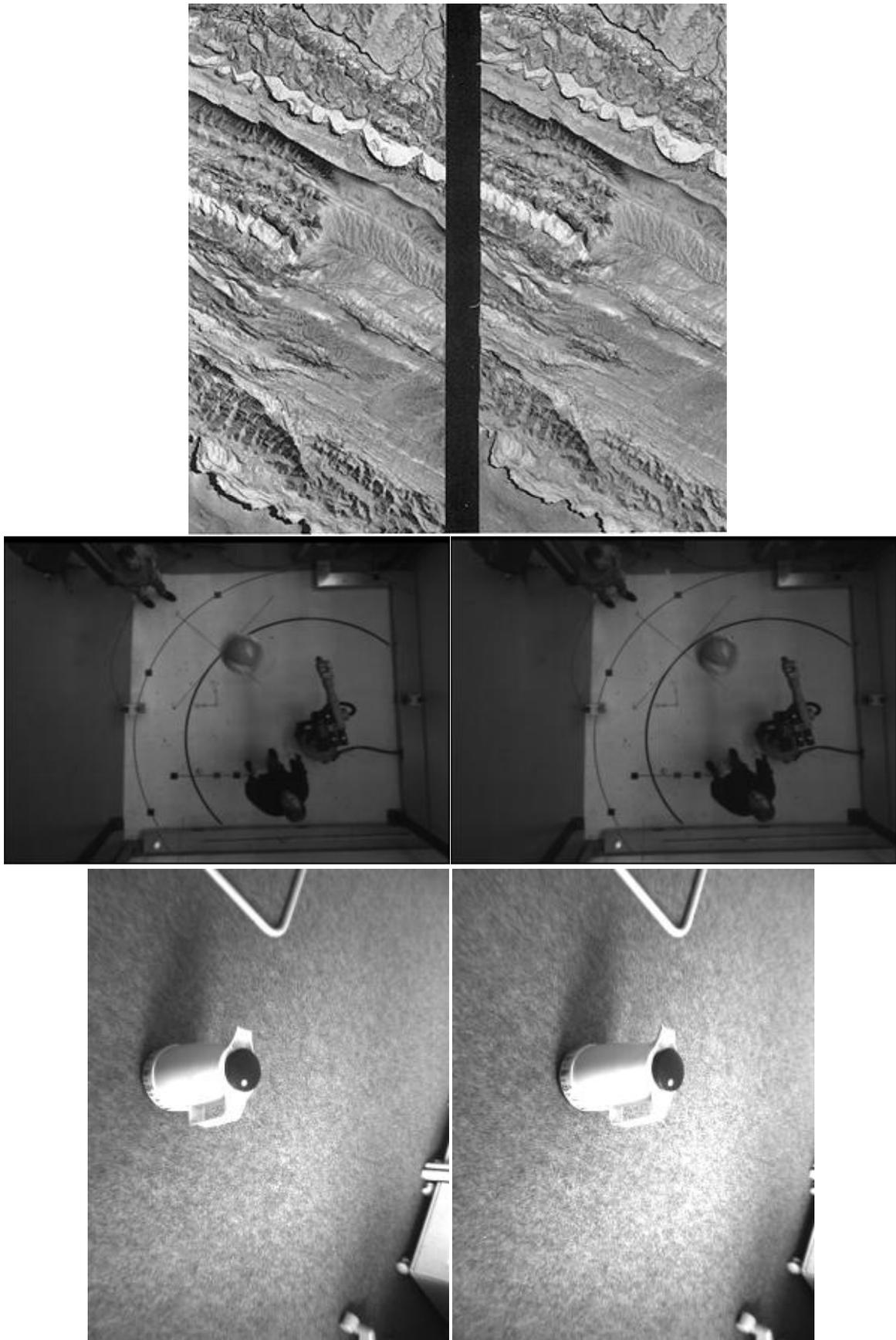


Рис. 3.1. Примеры стереоизображений с различным содержанием

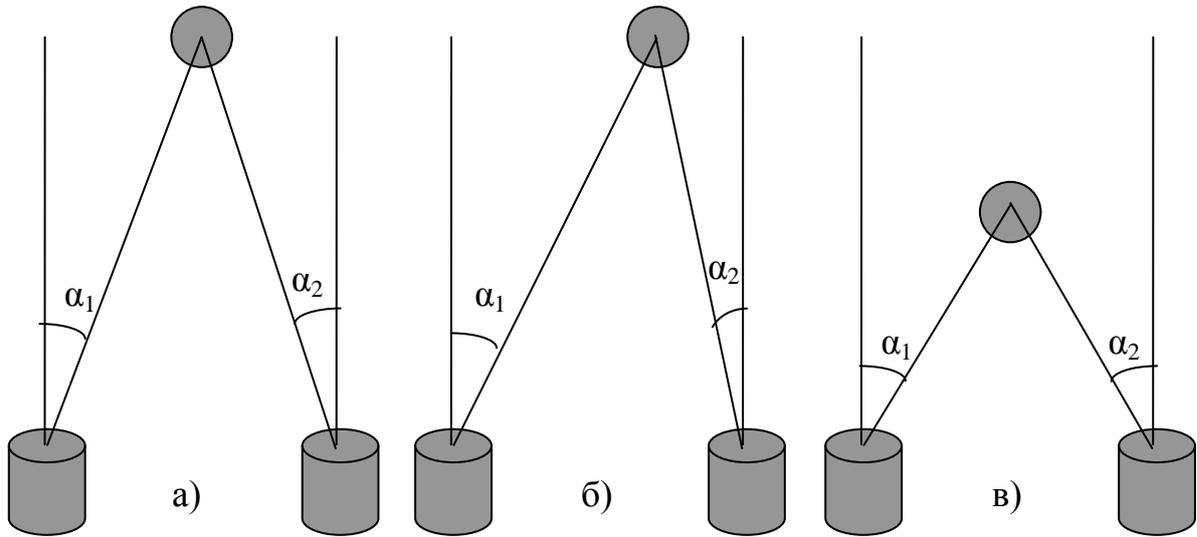


Рис. 3.2. Увеличение значения стереодиспаратности при уменьшении расстояния до объекта

Значение *стереодиспаратности* (разности углов, под которыми виден объект на двух камерах стереопары) при этом не зависит от положения объекта вдоль оси  $X$ , а зависит преимущественно от дальности. Действительно, разность  $\alpha_1 - \alpha_2$  одинакова для рис. 3.2а и 3.2б, в то время как для рис. 3.2в она существенно больше (следует иметь в виду, что  $\alpha_2 < 0$ ).

Значению стереодиспаратности, выраженному в угловой мере, соответствует значение стереодиспаратности, выраженное в единицах длины в плоскости изображения.

Запишем уравнение центральной проекции для первой камеры, считая, что начало координат привязано к ней:

$$x'/f' = x/z \text{ и } y'/f' = y/z.$$

Пусть смещение между камерами вдоль оси  $X$  равно  $d$ . Тогда уравнение центральной проекции для второй камеры (с учетом того, что начало координат для этой камеры смещено на величину  $-d$ ) будет

$$x''/f' = (x - d)/z \text{ и } y''/f' = y/z.$$

Получим разность координат одного и того же объекта в плоскости двух изображений:

$$\frac{x' - x''}{f'} = \frac{x - (x - d)}{z} \text{ и } \frac{y' - y''}{f'} = \frac{y - y}{z},$$

то есть

$$\Delta x = x' - x'' = \frac{df'}{z} \text{ и } \Delta y = y' - y'' = 0.$$

Можно начало координат поместить не в центр левой камеры, а в середину между камерами, однако основное уравнение, связывающее диспаратность с дальностью  $z$ , от этого не изменится

Таким образом, смещение объекта в плоскости изображения происходит вдоль той же оси, вдоль которой смещена камера. При этом диспаратность не

должна быть отрицательной, и она равна нулю при бесконечной (достаточно большой) удаленности объекта.

Координата  $z$ , как правило, априорно неизвестна. В связи с этим точке с координатами  $(x', y')$  может соответствовать любая точка, находящаяся на линии, задаваемой уравнениями  $x'' = x' + \frac{df'}{z}$  и  $y'' = y'$ . Эта линия называется *эпиполярной* и в случае задачи стереозрения представляет собой горизонтальную прямую (см. рис. 3.3). Как мы увидим позднее, при произвольном изменении ракурса камеры эпиполярная линия далеко не всегда является горизонтальной.

В результате, задача стереозрения сводится к отождествлению точек первого изображения с точками второго изображения при ограничении, заключающемся в том, что сопряженные точки лежат на соответствующих эпиполярных линиях. В результате такого отождествления при известной стереобазе и прочих параметрах камеры может быть восстановлено расстояние до объектов сцены.

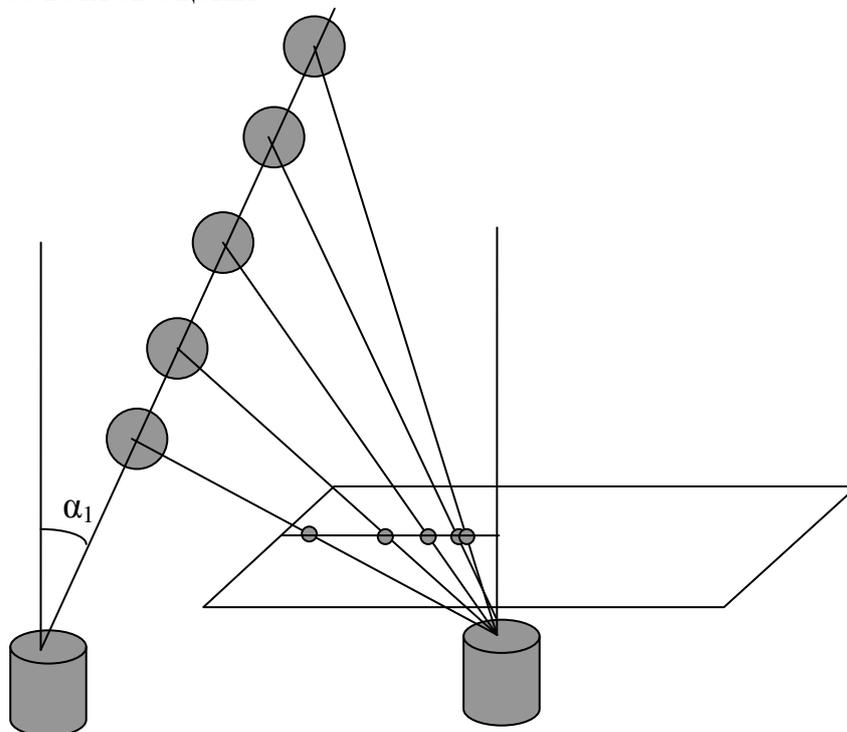


Рис. 3.3. Эпиполярная линия на втором изображении стереопары, задаваемая точкой на первом изображении

Как правило, значение диспаратности методами анализа изображений измеряется в пикселях. Для перевода координат  $(x', y')$  в пиксельные координаты необходимо знать параметры матрицы фотоприемника. Более доступной характеристикой, однако, является угол поля зрения камеры и число пикселей в формируемом ею изображении, что дает угловое разрешение камеры  $\theta_p$ , приходящееся на один пиксель.

К примеру, при поле зрения камеры, равном 30 градусов, и при числе пикселей на изображении вдоль оси  $X$ , равном 1024, угловое разрешение

составит примерно  $\theta_p = 5 \cdot 10^{-4}$  радиан/пиксель. Тогда при смещении в плоскости изображения  $\Delta x_p$ , равном 20 пикселей, угловое значение диспаратности составит 0.01 радиана. Эта величина приблизительно равна отношению  $d/z$ . При базе, равной 5 см, расстояние до объекта, которому соответствует это значение диспаратности, составит 5 м.

Таким образом, при малых угловых значениях диспаратности связь пиксельного смещения с расстоянием до объекта принимает простую форму

$$\Delta x_p \theta_p = d/z.$$

Как видно, зависимость смещения от дальности нелинейна. Учитывая то, что диспаратность измеряется с ненулевой погрешностью  $\varepsilon_{\Delta x}$  (как правило, составляющей 0.1–1 пиксель в зависимости от качества и содержания изображений), на разных расстояниях будет разная точность измерения глубины сцены: чем более далекими являются объекты, тем с меньшей точностью может быть измерена дальность до них. Поскольку при  $z = \infty$  диспаратность  $\Delta x_p = 0$ , объекты, для которых может быть измерена дальность, должны находиться на расстоянии

$$z < d/(\varepsilon_{\Delta x} \theta_p).$$

Для приведенного выше примера ( $\theta_p = 5 \cdot 10^{-4}$  и  $d=5$  см) при  $\varepsilon_{\Delta x} = 1$  получим  $z < 100$  м.

Теперь посмотрим, как именно зависит точность измерения диспаратности от расстояния  $z$ . Запишем уравнения диспаратности для некоторого смещения  $\Delta x_p$ , соответствующего расстоянию  $z$ , и для смещения  $\Delta x_p + \varepsilon_{\Delta x}$ , соответствующего расстоянию  $z - \varepsilon_z$ , где  $\varepsilon_z$  – ошибка измерения дальности, соответствующая ошибке измерения диспаратности  $\varepsilon_{\Delta x}$ :

$$\Delta x_p \theta_p = d/z \text{ и } (\Delta x_p + \varepsilon_{\Delta x}) \theta_p = d/(z - \varepsilon_z).$$

Подставляя  $\Delta x_p$  из первого уравнения во второе получаем

$$(z - \varepsilon_z) \theta_p = \frac{d}{\frac{d}{\theta_p z} + \varepsilon_{\Delta x}} \Rightarrow z - \varepsilon_z = \frac{dz}{d + \varepsilon_{\Delta x} \theta_p z} \Rightarrow \varepsilon_z = \frac{z(d + \varepsilon_{\Delta x} \theta_p z) - dz}{d + \varepsilon_{\Delta x} \theta_p z} \Rightarrow$$

$$\Rightarrow \varepsilon_z = \frac{\varepsilon_{\Delta x} \theta_p z^2}{d + \varepsilon_{\Delta x} \theta_p z}.$$

При  $d \gg \varepsilon_{\Delta x} \theta_p z$  (это условие, как правило, выполняется для значений  $z$ , при которых может быть измерена диспаратность) получаем

$$\varepsilon_z \approx \varepsilon_{\Delta x} \theta_p z^2 / d. \quad (3.1)$$

Может показаться удивительным, что ошибка измерения расстояния пропорциональна  $z^2$ , а не  $z$ . В действительности, если представить стереокамеру с некоторой базой, наблюдающей объект на некотором

расстоянии, и произвести увеличение в одинаковое число раз всех расстояний, то пиксельное смещение на изображении не изменится.

Иными словами, если для стереокамеры с базой 1 см ошибка измерения диспаратности на расстоянии 1 м составляет 10 см, то у этой же камеры с базой 1 км ошибка измерения диспаратности на расстоянии 100 км составит 10 км. То есть при фиксированном отношении  $z/d$  ошибка увеличивается пропорционально  $z$ , исходя из простого условия подобия. Это и приводит к тому, что при фиксированной базе ошибка возрастает пропорционально  $z^2$ .

Для приведенного выше примера ( $\theta_p = 5 \cdot 10^{-4}$  и  $d=5$  см,  $\varepsilon_{\Delta x} = 1$ ) точность измерения глубины на  $z=1$  м составит 1 см, а на  $z=10$  м составит 1 м. В связи с этим при использовании метода стереозрения в задаче восстановления расстояний необходимо выбирать стереобазу и разрешение камеры таким образом, чтобы для характерной удаленности объектов дальность могла восстанавливаться с приемлемой точностью.

Рассмотрим, к примеру, возможность определения высоты зданий со спутника, пролетающего на высоте 50 км и снимающего один и тот же участок местности с разных точек орбиты (при известной ориентации спутника эту задачу можно свести к задаче стереозрения). Насколько удаленными должны быть точки съемки, чтобы погрешность определения высоты составила 1 метр при  $\theta_p = 5 \cdot 10^{-4}$  и  $\varepsilon_{\Delta x} = 1$ ?

Из уравнения (3.1) получаем соотношение  $d \approx \varepsilon_{\Delta x} \theta_p z^2 / \varepsilon_z$ , откуда  $d \approx 1250$  км. Очевидно, данный результат говорит о том, что спутник с указанными характеристиками не может быть использован для решения поставленной задачи, и необходимо либо увеличивать разрешение камеры (действительно, использование камеры, имеющей матрицу шириной всего 1024 пикселя, вряд ли оправдано), либо понижение орбиты спутника, либо улучшение алгоритма измерения стереодиспаратности.

Таким же образом можно решать вопросы о том, может ли использоваться стереомикроскоп для определения дефектов печатных плат или стереокамера для определения неровностей поверхности при посадке вертолета в скалистой местности. Существует много задач, в которых стереокамеры позволяют получить достаточную точность определения дальности, однако прежде, чем предлагать решение некоторой задачи с использованием стереокамеры, необходимо убедиться, что требуемая точность принципиально достижима.

Для реализации стереозрения необходимо выполнение сопоставление изображений, полученных с разных ракурсов, то есть необходимо выполнять отождествление сопряженных точек.

### *Корреляционные методы стереозрения*

Задача стереозрения является одним из непосредственных приложений метода локальной корреляции, поскольку в изображениях стереопары отсутствуют глобальные геометрические искажения, а имеются только

локальные сдвиги. Однако реализация метода локальной корреляции применительно к этой задаче может быть весьма разнообразной.

Поскольку сдвиги точек происходят вдоль горизонтальных эпиполярных линий, нет смысла вычислять двумерное кросскорреляционное поле. В связи с этим вместо вычисления кросскорреляционного поля через преобразование Фурье в данной задаче обычно используется непосредственное вычисление значения корреляции при разных сдвигах по оси  $X$

$$C_{f_1, f_2}(\Delta x) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f_1(x, y) f_2(x - \Delta x, y),$$

причем часто используется не квадратное окно.

Здесь оправданным является использование не значения корреляции, а коэффициента корреляции, для получения которого необходимо выполнить нормировку

$$c_{f_1, f_2}(\Delta x) = \frac{C_{f_1, f_2}(\Delta x)}{\frac{1}{MN} \left[ \left( \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f_1^2(x, y) \right) \left( \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f_2^2(x - \Delta x, y) \right) \right]^{0.5}}.$$

Алгоритм стереозрения при этом весьма простой. Вокруг каждой точки первого изображения берется окно размера  $M \times N$  и для разных значений  $\Delta x \in [0, \Delta x_{\max}]$  вычисляется коэффициент корреляции, и определяется значение сдвига, при котором коэффициент максимален. Этот сдвиг и является определенной диспаратностью для данной точки.

На рис. 3.4 представлена стереопара, для которой построено поле стереодиспаратности при разных параметрах алгоритма. На рис. 3.4в значение диспаратности отображено для всех точек. Видно, что в точках, находящихся внутри зон однородной яркости диспаратности вычисляются неверно. На рис. 3.4г представлено то же поле диспаратности, но только для точек, в которых коэффициент корреляции больше 0.9. Это позволяет достаточно надежно выделить точки с корректно измеренной диспаратностью.

На рис. 3.5 представлены еще два примера вычисления стереодиспаратности (для первого изображения поле диспаратности показано без фильтрации по величине коэффициента корреляции, а для второго изображения – с фильтрацией). Как видно, корреляционный подход в данном случае может давать удовлетворительный результат.

Основные проблемы в смысле качества решения у данного подхода связаны с наличием шумов на поле диспаратности и с большим его сглаживанием, причем попытка уменьшить шумы за счет увеличения размеров окна корреляции ведет к увеличению сглаживания и наоборот (см. рис. 3.4д,е). Кроме того, корреляционный метод требует большого объема вычислений и работает недостаточно быстро. Уменьшение размера окна позволяет получить решения, работающие в реальном масштабе времени, однако шумы при этом увеличиваются катастрофически.

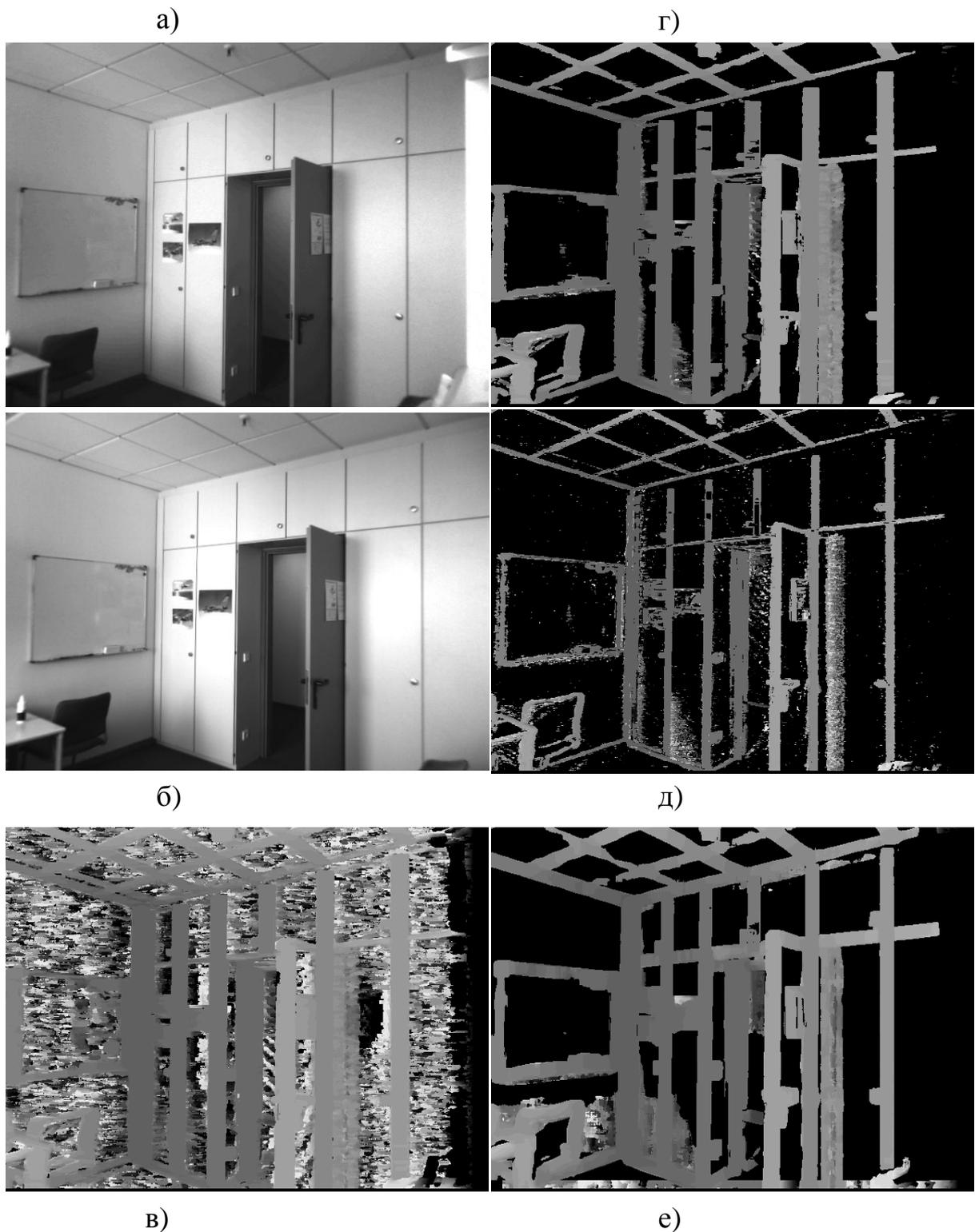


Рис. 3.4. Пример стереопары (а, б) и результаты корреляционного вычисления значений стереодиспаратности: для окна 20x4 без исключения точек (в); для окон 20x4, 10x2 и 16x16 (г, д, е соответственно) с исключением точек с низким коэффициентом корреляции

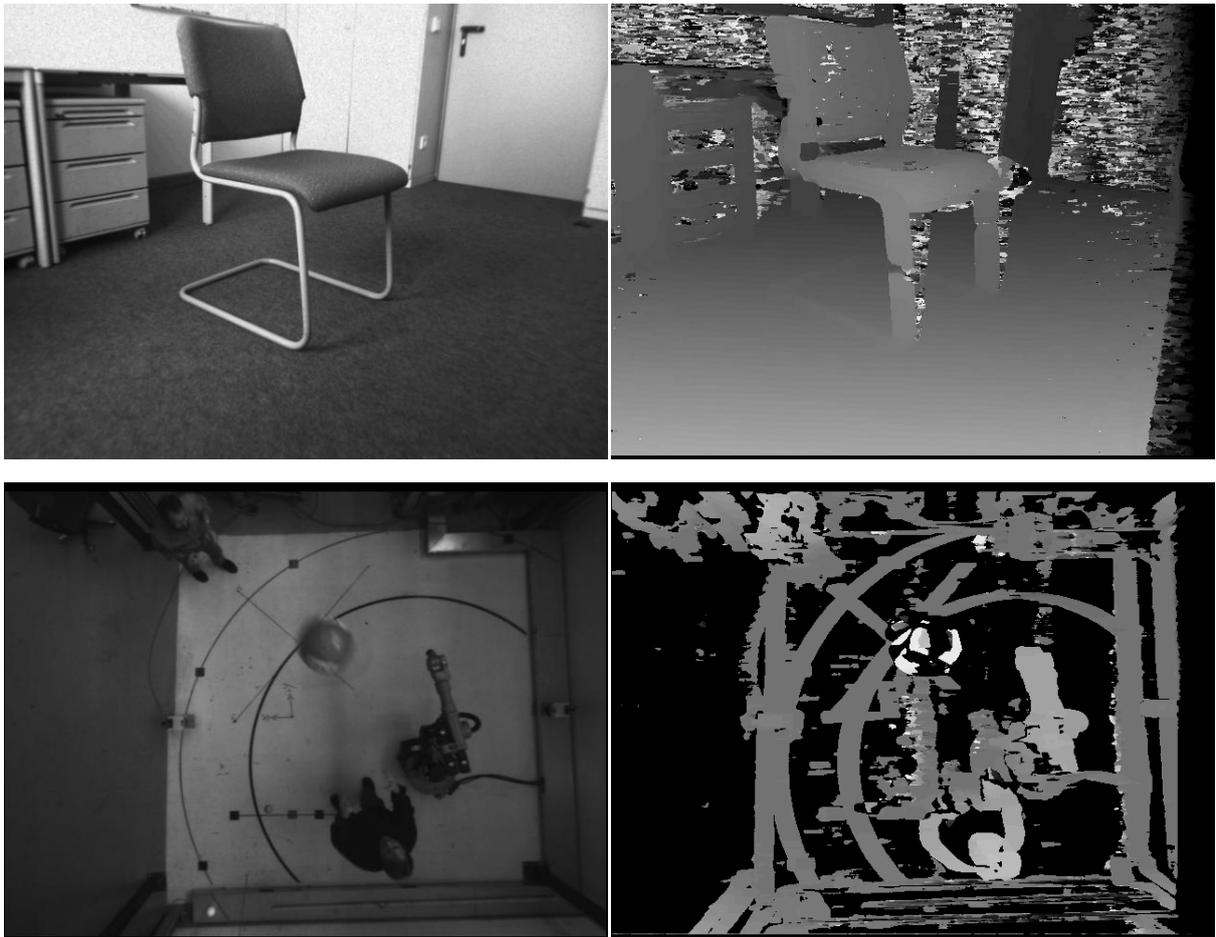


Рис. 3.5. Примеры построения карт стереодиспаратности корреляционным методом

Можно заметить, что наиболее надежное вычисление диспаратности происходит вблизи контуров. При этом, однако, в местах разрывов по глубине корреляционный отклик может нарушаться. Это приводит к идее использования контурного подхода к стереозрению, в рамках которого значения диспаратности также будут измеряться для точек контура, и при этом трудности, связанные с разрывами по глубине будут менее существенными.

#### *Контурные и структурные методы стереозрения*

Общей идеей контурных методов стереозрения является определение стереодиспаратности для контурных точек с учетом их принадлежности одним и тем же цепочкам. Для каждой цепочки или даже сегмента контура может производиться независимый поиск смещения. При этом для второго изображения выполняется преобразование расстояния, и для каждого сегмента контуров первого изображения находится собственное смещение путем минимизации меры Хаусдорфа (или оптимизации некоторой аналогичной меры, основанной на функции расстояния).

На рис. 3.6 представлен пример непосредственного наложения контуров изображения, представленного на рис. 3.4а, на функцию  $\frac{s}{s + D_{C_2}(x, y)}$ ,

примененную к контурам изображения 3.4б ( $D_{C_2}(x, y)$ ) – преобразование расстояния для контуров  $C_2$ ). Путем применения чамферного сопоставления (2.22) с использованием только горизонтальных сдвигов для каждой цепочки контуров можно получить для них индивидуальные значения диспаратностей (рис. 3.7).

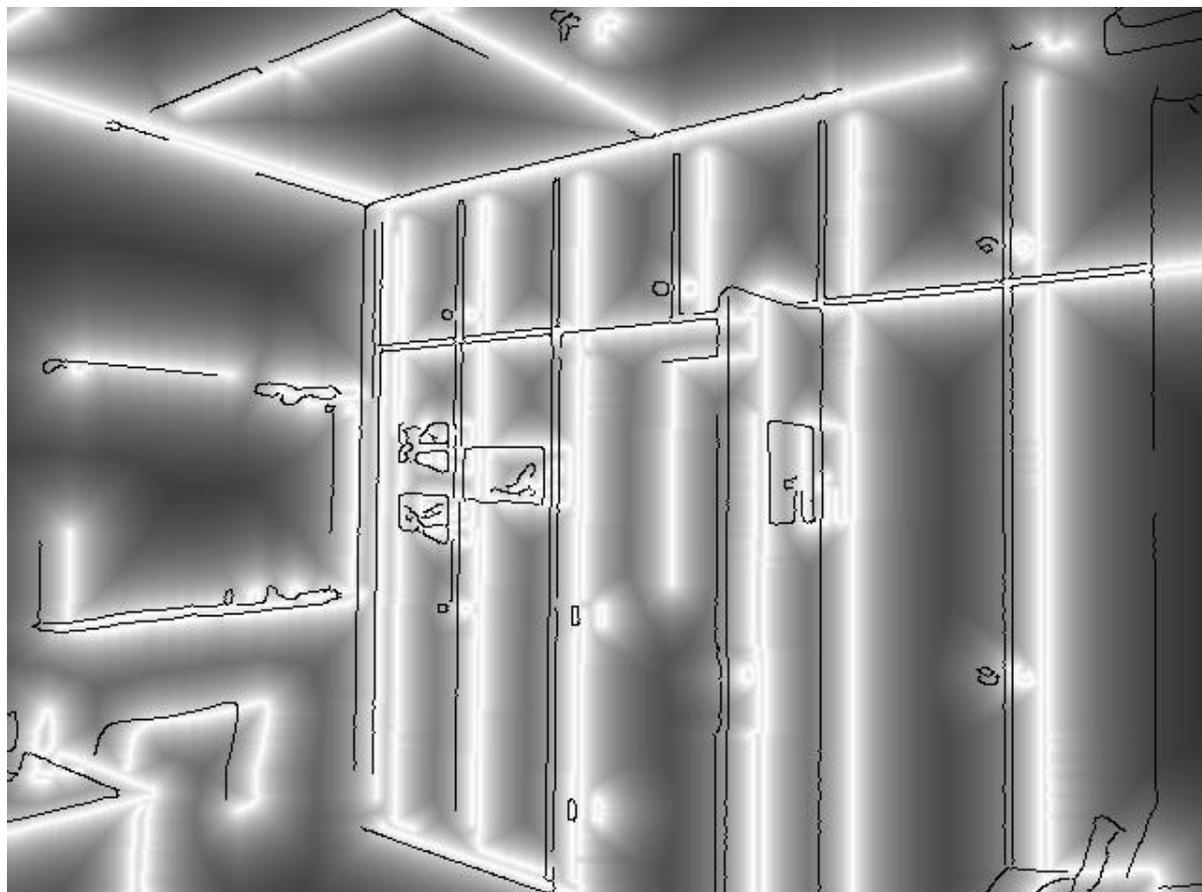


Рис. 3.6. Наложение контуров первого изображения на функцию дальности, полученную на основе преобразования расстояния, примененного к контурам второго изображения

Следует, однако, иметь в виду, что только для вертикальных контуров оценка величины стереодиспаратности будет надежной. Кроме того, поиск смещения для отдельных цепочек контуров может приводить к неоднозначности (особенно, в данном случае, когда вертикальные цепочки контуров мало отличимы друг от друга). Если же точки контура имеют разную глубину, то на двух изображениях стереопары этот контур будет иметь разный наклон, что приведет к увеличению меры Хаусдорфа и снижению вероятности определения правильного смещения.

Частично эти трудности могут быть преодолены в рамках структурного подхода, в котором сопоставляются не сами контуры, а только некоторые их характерные точки. К примеру, на рис. 3.8 представлено структурное описание изображения стереопары, приведенного на рис. 3.5. На этом рисунке отображены контурные точки максимальной кривизны, соединенные прямыми линиями.

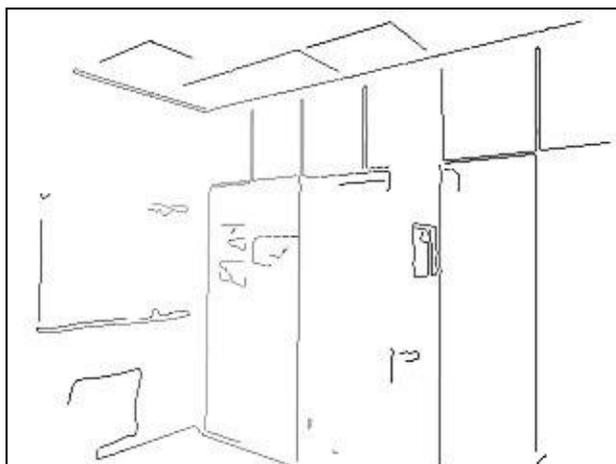


Рис. 3.7. Результат вычисления диспаратности для контурных точек стереопары 3.4а, б

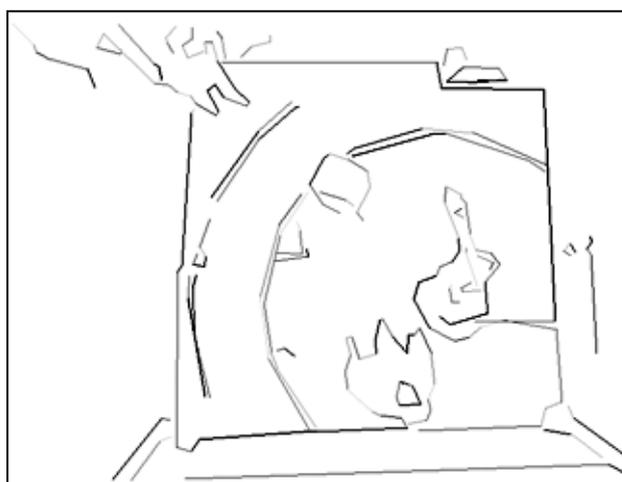


Рис. 3.8. Пример построения структурного описания стереоизображения стереопары

В этом случае сопоставляются только точки максимальной кривизны, для каждой из которых может быть выдвинуто несколько гипотез сопоставления (рис. 3.9). Далее из всех гипотез, выдвинутых для каждой характерной точки, выбираются взаимно непротиворечивые гипотезы (совместная проверка гипотез является одной из особенностей структурного подхода), что обеспечивает снятие неопределенности, возникающей в процессе сопоставления. Кроме того, характерным точкам, располагающимся на одном контуре, может быть разрешено иметь не совпадающие, но не сильно отличающиеся смещения.

Для характерных точек, располагающихся на горизонтальных участках контуров, диспаратности могут быть экстраполированы с соседних характерных точек. После определения диспаратности для характерных точек, значения диспаратности могут быть определены для всех точек контуров путем их эластичного сопоставления (см. рис. 3.10). Для распространения значений диспаратностей на все остальные точки требуются дополнительные методы.

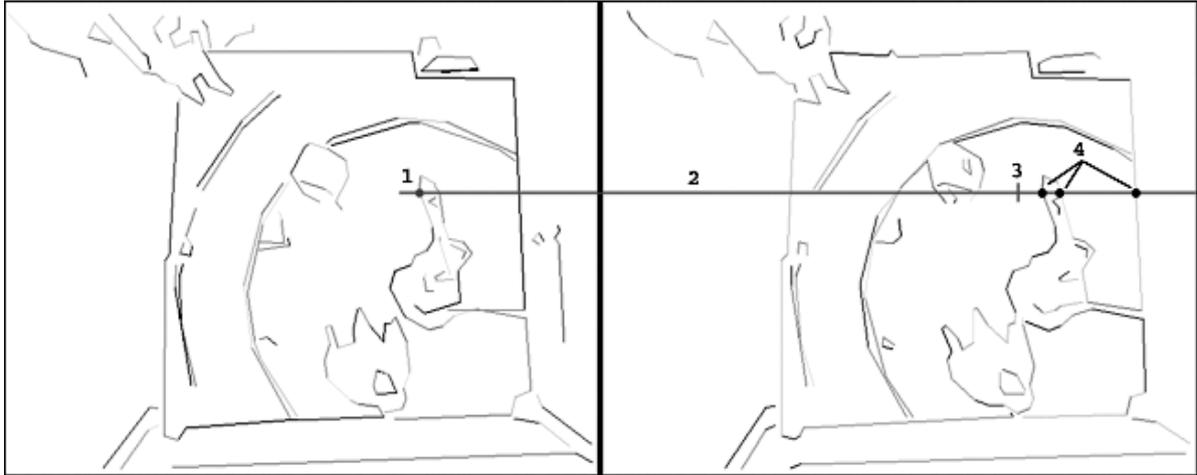


Рис. 3.9. Формирование для точки левого изображения (1) на основе эпиллярной линии (2) гипотез сопоставления (4) справа от точки нулевой диспаратности (3)



Рис. 3.10. Пример восстановления глубины сцены для контурных точек структурным методом

### Вопросы и упражнения

1. В чем заключается идея стереозрения и для решения каких проблем оно может применяться?
2. Как зависит точность определения глубины сцены в зависимости от стереобазы и расстояний до объектов?

3. Что такое эпиполярная линия? Какой вид эпиполярные линии имеют в случае задачи стереозрения?
4. Как применяются корреляционные методы для решения задачи стереозрения? В чем их ограничение?
5. Зачем необходимо нормировать значение корреляции при вычислении стереодиспаратности?
6. Каким образом используется мера Хаусдорфа для решения задачи стереозрения?
7. В чем преимущество контурных методов стереозрения по сравнению с корреляционными?
8. Какова идея структурных методов стереозрения? Чем они отличаются от контурных методов?
9. В чем существенное отличие при использовании структурных методов для сопоставления изображений в случае глобального пространственного преобразования изображений и при сопоставлении изображений стереопары?

### 3.2. Сопоставление изображений трехмерных сцен

#### *Относительная ориентация камер*

Ранее мы рассмотрели две задачи: сопоставление изображений плоских сцен и стереозрение. В первом случае координаты сопряженных точек на изображениях оказывались связанными единым глобальным пространственным преобразованием. Во втором случае взаимное расположение камер считалось известным (причем рассматривался простейший случай сдвига вдоль одной оси), что позволяло рассматривать только проблему определения дальности до точек сцены. В общем случае не известны ни параметры взаимного расположения камер, ни дальность до точек сцены. Возникает вопрос, можно ли в этом случае одновременно оценить как дальности до точек сцены, так и взаимное расположение камер.

Для начала посмотрим, как можно связать системы координат двух камер, обладая информацией не об абсолютных координатах сопряженных точек  $\mathbf{r}_1 = (x_1, y_1, z_1)$  и  $\mathbf{r}_2 = (x_2, y_2, z_2)$ , а лишь об их проекциях на плоскости изображений  $(x'_1, y'_1)$  и  $(x'_2, y'_2)$ . Как мы выяснили ранее, величины  $z_1$  и  $z_2$  из уравнения этой связи исключить невозможно, поэтому будем рассматривать их как дополнительные параметры.

Осуществим шаги по исключению физических координат точек из уравнений центральной проекции. Запишем в развернутой форме уравнение (2.4) связи систем координат двух камер  $\mathbf{r}_2 = \mathbf{R}\mathbf{r}_1 + \mathbf{r}_o$  (уравнение абсолютной ориентации)

$$x_2 = r_{11}x_1 + r_{12}y_1 + r_{13}z_1 + \Delta x,$$

$$y_2 = r_{21}x_1 + r_{22}y_1 + r_{23}z_1 + \Delta y.$$

$$z_2 = r_{31}x_1 + r_{32}y_1 + r_{33}z_1 + \Delta z.$$

Воспользуемся уравнением центральной проекции (2.1) для первой и второй камер

$$\begin{aligned}x_1 &= z_1 x'_1 / f'_1, \quad y_1 = z_1 y'_1 / f'_1, \\x_2 &= z_2 x'_2 / f'_2, \quad y_2 = z_2 y'_2 / f'_2,\end{aligned}$$

и исключим из уравнений связи пространственные координаты:

$$\begin{aligned}z_2 x'_2 / f'_2 &= r_{11}(z_1 x'_1 / f'_1) + r_{12}(z_1 y'_1 / f'_1) + r_{13}z_1 + \Delta x \\z_2 y'_2 / f'_2 &= r_{21}(z_1 x'_1 / f'_1) + r_{22}(z_1 y'_1 / f'_1) + r_{23}z_1 + \Delta y \\z_2 &= r_{31}(z_1 x'_1 / f'_1) + r_{32}(z_1 y'_1 / f'_1) + r_{33}z_1 + \Delta z\end{aligned}$$

Перепишем эти уравнения в форме

$$\begin{aligned}(z_2 / z_1)(f'_1 / f'_2)x'_2 &= r_{11}x'_1 + r_{12}y'_1 + r_{13}f'_1 + f'_1\Delta x / z_1 \\(z_2 / z_1)(f'_1 / f'_2)y'_2 &= r_{21}x'_1 + r_{22}y'_1 + r_{23}f'_1 + f'_1\Delta y / z_1 \\(z_2 / z_1)f'_1 &= r_{31}x'_1 + r_{32}y'_1 + r_{33}f'_1 + f'_1\Delta z / z_1\end{aligned}\tag{3.2}$$

Эти уравнения называются *уравнениями относительной ориентации*.

Если известны коэффициенты матрицы вращения  $\mathbf{R}$  и вектора смещения  $\mathbf{r}_o$ , то из уравнений (3.2) несложно получить дальности  $z_1$  и  $z_2$  для каждой пары сопряженных точек. Это будет соответствовать общему случаю задачи стереозрения, когда камеры ориентированы произвольным, но известным образом. Эта задача мало отличается от задачи обычного стереозрения, поэтому здесь мы ее не рассматриваем. Гораздо сложнее и интереснее случай, когда величины  $\mathbf{R}$  и  $\mathbf{r}_o$  в уравнениях (3.2) неизвестны.

В этих уравнениях присутствует 14 неизвестных: 9 параметров матрицы  $\mathbf{R}$ , 3 параметра вектора смещения  $\mathbf{r}_o$ , а также величины  $z_1$  и  $z_2$ . При этом, однако, все величины  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  и  $z_2$  в уравнениях всегда делятся на  $z_1$ . Это говорит о том, что однозначного решения получено быть не может: если  $\Delta x^{(0)}$ ,  $\Delta y^{(0)}$ ,  $\Delta z^{(0)}$ ,  $z_1^{(0)}$  и  $z_2^{(0)}$  – решение, то и  $a\Delta x^{(0)}$ ,  $a\Delta y^{(0)}$ ,  $a\Delta z^{(0)}$ ,  $az_1^{(0)}$  и  $az_2^{(0)}$  является решением при любом значении  $a$ . Иными словами, все длины и расстояния определяются с точностью до масштаба (то есть определяются относительные, а не абсолютные длины).

Поскольку решений получается бесконечно много, необходимо накладывать дополнительное ограничение на указанные 5 параметров (к примеру, считать вектор смещения единичным).

Кроме указанного ограничения (уменьшающего число независимых переменных на 1) можно также воспользоваться условием ортогональности матрицы  $\mathbf{R}$  (что задает 6 дополнительных ограничений, накладываемых на переменные). Если же фокусное расстояние для второго снимка отличается от фокусного расстояния для первого снимка и является неизвестным, то число неизвестных следует увеличить на 1.

Как видно, имея одну пару сопряженных точек, дающую 3 уравнения и два неизвестных, невозможно определить относительную ориентацию, задаваемую двенадцатью неизвестными (при семи дополнительных ограничениях). Каждая пара сопряженных точек дает 2 новые неизвестные и 3 новых уравнения. Тогда при  $n$  отождествленных сопряженных точек будет

$12 + 2n$  неизвестных и  $7 + 3n$  связывающих их уравнений. Теоретически минимальное число пар точек, необходимое для определения параметров относительной ориентации, равно 5.

Решив систему уравнений относительной ориентации при заданных парах сопряженных точек, несложно получить для любой пары сопряженных точек значения дальностей, поделив первое уравнение системы (3.2) на второе:

$$\frac{x'_2}{y'_2} = \frac{r_{11}x'_1 + r_{12}y'_1 + r_{13}f'_1 + f'_1\Delta x / z_1}{r_{21}x'_1 + r_{22}y'_1 + r_{23}f'_1 + f'_1\Delta y / z_1},$$

откуда

$$\frac{1}{z_1} = \frac{x'_2(r_{21}x'_1 + r_{22}y'_1 + r_{23}f'_1) - y'_2(r_{11}x'_1 + r_{12}y'_1 + r_{13}f'_1)}{y'_2 f'_1 \Delta x - x'_2 f'_1 \Delta y}.$$

В случае отсутствия перемещения в плоскости, перпендикулярной оптической оси (при равенстве нулю величин  $\Delta x$ ,  $\Delta y$ ), для получения расстояний следует воспользоваться также третьим уравнением системы (3.2).

Однако даже если необходимое число пар отождествленных сопряженных точек известно, следует иметь в виду, что решение уравнений (3.2) не дает абсолютных значений параметров  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ . Все они определяются с точностью до постоянного множителя. Поскольку система (3.2) решается одновременно для всех пар точек, и для каждой пары точек значения  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  должны быть одинаковы, значения дальностей всех точек находятся с точностью до одного и того же множителя (то есть определяется относительная удаленность точек сцены).

Чтобы восстановить абсолютные дальности до точек сцены, одной только информации о сопряженных точках на двух изображениях недостаточно. Для этого необходимо либо знать модуль вектора переноса (как, например, в случае задачи стереозрения, в которой была известна стереобаза), либо иметь какой-то нормировочный объект с известным размером или расстоянием до него. Подобные опорные объекты могут использовать не только для восстановления абсолютного масштаба, но также и для отождествления сопряженных точек, о чем будет сказано несколько позднее.

Следует отметить, что взаимная ориентация камер восстанавливается однозначно в отличие от вектора переноса.

Уравнения относительной ориентации (3.2), к сожалению, нелинейны. Кроме того, их количество растет с числом пар сопряженных точек, в результате чего решение соответствующей системы уравнений оказывается весьма проблематичным. Вместо этого может использоваться следующий более практичный подход.

### Эпиполярная геометрия в случае двух произвольных проекций

Как мы видели раньше, в случае двух произвольных проекций для объемной сцены (точки которой не лежат в одной плоскости) координаты сопряженных точек на двух изображениях связаны через расстояние до этих точек. Таким образом, если мы задаем некоторую точку на первом изображении, сопряженная ей точка на втором изображении должна лежать на некоторой линии, называемой эпиполярной. К примеру, для изображений стереопары эпиполярная линия для точки является горизонтальной линией, ордината которой совпадает с ординатой точки. Выведем уравнение эпиполярной линии для случая двух произвольных проекций путем исключения глубин точки.

Для исключения глубины точки с помощью третьего уравнения системы (3.2) выразим отношение

$$(z_2 / z_1) = (r_{31}x'_1 + r_{32}y'_1 + r_{33}f'_1 + f'_1\Delta z / z_1) / f'_1$$

и подставим его в первые два уравнения этой же системы:

$$(r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33} + \Delta z / z_1) x'_2 f'_1 / f'_2 = r_{11}x'_1 + r_{12}y'_1 + r_{13}f'_1 + \Delta x f'_1 / z_1,$$

$$(r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33} + \Delta z / z_1) y'_2 f'_1 / f'_2 = r_{21}x'_1 + r_{22}y'_1 + r_{23}f'_1 + \Delta y f'_1 / z_1.$$

При заданных положениях камеры и координатах точки на первом изображении  $(x'_1, y'_1)$  данные уравнения являются параметрическим заданием линии на втором изображении. В качестве параметра выступает величина  $z_1$ . Исключим ее.

$$\begin{aligned} (\Delta z x'_2 f'_1 / f'_2 - \Delta x f'_1) / z_1 &= \\ &= r_{11}x'_1 + r_{12}y'_1 + r_{13}f'_1 - (r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33}) x'_2 f'_1 / f'_2, \\ (\Delta z y'_2 f'_1 / f'_2 - \Delta y f'_1) / z_1 &= \\ &= r_{21}x'_1 + r_{22}y'_1 + r_{23}f'_1 - (r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33}) y'_2 f'_1 / f'_2. \end{aligned}$$

Откуда получаем уравнение, задающее эпиполярную линию.

$$\begin{aligned} (\Delta z y'_2 f'_1 / f'_2 - \Delta y f'_1)(r_{11}x'_1 + r_{12}y'_1 + r_{13}f'_1 - (r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33}) x'_2 f'_1 / f'_2) = \\ = (\Delta z x'_2 f'_1 / f'_2 - \Delta x f'_1)(r_{21}x'_1 + r_{22}y'_1 + r_{23}f'_1 - (r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33}) y'_2 f'_1 / f'_2). \end{aligned} \quad (3.3)$$

Отметим, что единственные нелинейные (по  $x'_2$  и  $y'_2$ ) члены в обеих частях уравнения одинаковы  $(r_{31}x'_1 / f'_1 + r_{32}y'_1 / f'_1 + r_{33})\Delta z x'_2 y'_2 (f'_1 / f'_2)^2$  и, следовательно, сокращаются. То есть это уравнение задает прямую линию.

Это уравнение можно переписать в более компактной форме:

$$(x'_2, y'_2, 1)\mathbf{F}(x'_1, y'_1, 1)^T = 0, \quad (3.4)$$

где  $\mathbf{F}$  – так называемая *фундаментальная матрица*.

$$\begin{aligned} (x'_2 \quad y'_2 \quad 1) \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} x'_1 \\ y'_1 \\ 1 \end{pmatrix} = \\ = x'_2(f_{11}x'_1 + f_{12}y'_1 + f_{13}) + y'_2(f_{21}x'_1 + f_{22}y'_1 + f_{23}) + (f_{31}x'_1 + f_{32}y'_1 + f_{33}) = 0. \end{aligned}$$

Чтобы получить уравнение (3.4) из уравнения (3.3), нужно определить коэффициенты фундаментальной матрицы следующим образом:

$$\begin{aligned} f_{11} &= (\Delta y_{r31} - \Delta z_{r21})a & f_{12} &= (\Delta y_{r32} - \Delta z_{r22})a & f_{13} &= (\Delta y_{r33} - \Delta z_{r23})af'_1 \\ f_{21} &= (\Delta z_{r11} - \Delta x_{r31})a & f_{22} &= (\Delta z_{r12} - \Delta x_{r32})a & f_{23} &= (\Delta z_{r13} - \Delta x_{r33})af'_1 \\ f_{31} &= (\Delta x_{r21} - \Delta y_{r11})f'_1 & f_{32} &= (\Delta x_{r22} - \Delta y_{r12})f'_1 & f_{33} &= (\Delta x_{r23} - \Delta y_{r13})f'_1 f'_1 \end{aligned}$$

где  $a = f'_1 / f'_2$ .

Если известна фундаментальная матрица, то не составляет трудности для любой точки определить эпиполярную линию, которую эта точка задает на другом изображении. Преимущество уравнения (3.4), связывающего проекции точки через фундаментальную матрицу, по сравнению с уравнениями (3.2), связывающими проекции точки через матрицу вращения камеры и вектор смещения, в том, что уравнения (3.4) линейны по параметрам (коэффициентам матрицы  $\mathbf{F}$ ). При заданном наборе сопряженных точек фундаментальная матрица может быть восстановлена достаточно просто с помощью метода наименьших квадратов.

Стоит отметить, что фундаментальная матрица восстанавливается с точностью до постоянного множителя. Действительно, если матрица  $\mathbf{F}$  удовлетворяет уравнению (3.4), то и матрица  $\alpha\mathbf{F}$  также удовлетворяет этому уравнению (что аналогично восстановлению вектора переноса и дальностей с точностью до множителя). При использовании метода наименьших квадратов оказывается необходимым накладывать некоторое дополнительное условие, чтобы получить однозначное (и невырожденное) решение. Обычно это делается одним из двух путей: либо элемент  $f_{33}$ , либо определитель матрицы полагается равным 1 (последнее более сложно в реализации).

К примеру, точки, представленные на рис. 3.11а задают фундаментальную матрицу (в случае, когда в качестве нормировки используется приравнивание единице элемента  $f_{33}$ ).

$$\begin{pmatrix} 2.22 \cdot 10^{-5} & 1.48 \cdot 10^{-4} & 3.33 \cdot 10^{-3} \\ -1.56 \cdot 10^{-4} & 3.42 \cdot 10^{-5} & 3.24 \cdot 10^{-2} \\ -1.33 \cdot 10^{-2} & -3.11 \cdot 10^{-2} & 1.0 \end{pmatrix}$$

и систему эпиполярных линий, представленных на рис. 3.11б.

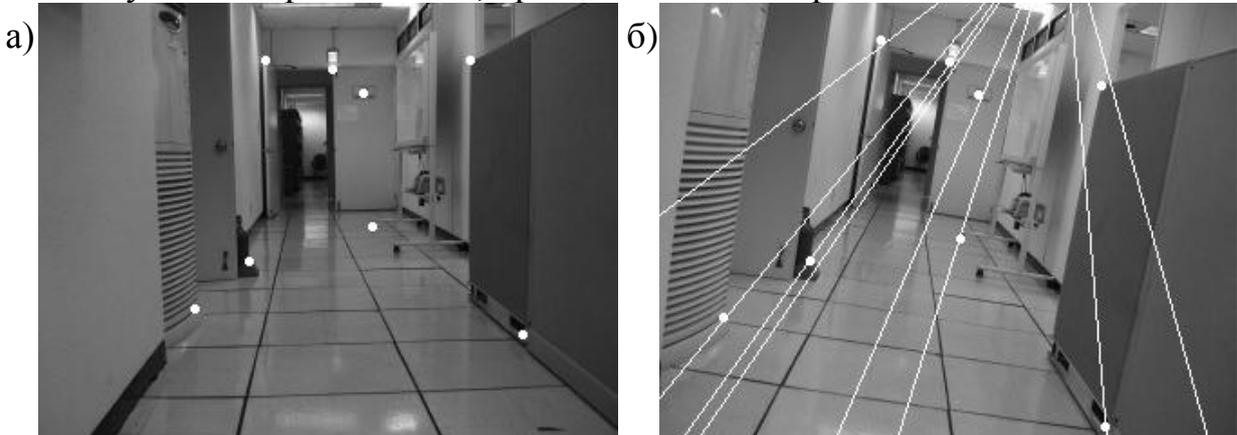


Рис. 3.11. Структура эпиполярных линий для двух произвольных проекций

Задачи определения параметров абсолютной или относительной ориентации камеры по известным сопряженным точкам относятся к задачам фотограмметрии.

В компьютерном зрении решается задача сопоставления изображений трехмерных сцен, в которой сопряженные точки заранее неизвестны, и их необходимо отождествить. При этом все сопряженные точки должны удовлетворять одной и той же (неизвестной) относительной ориентации камер. При решении проблемы одновременного отождествления сопряженных точек и установления относительной ориентации камер можно также идти двумя путями.

1. Производить поиск матрицы вращения  $\mathbf{R}$  и вектора переноса  $\mathbf{r}_o$  совместно с определением дальности до точек, используя уравнения (3.2). Как отмечалось выше, при этом требуется не менее 5 пар отождествленных точек, и система уравнений получается существенно нелинейной.
2. Производить поиск фундаментальной матрицы  $\mathbf{F}$ , то есть работать полностью в плоскостях изображений без определения дальности. Каждая пара сопряженных точек добавляет одно уравнение  $(x'_2, y'_2, 1)\mathbf{F}(x'_1, y'_1, 1)^T = 0$  и не добавляет неизвестных (в отличие от предыдущего случая, когда каждая пара добавляла 3 уравнения и 2 неизвестных). Матрица определяется с точностью до постоянного множителя, поэтому имеется всего 8 неизвестных. Кроме того, коэффициенты фундаментальной матрицы также не являются полностью независимыми. В связи с этим нужно лишь семь пар сопряженных точек для их определения (при этом, однако, решение оказывается неоднозначным, и для устранения этой неоднозначности нужна восьмая точка).

Таким образом, хотя подход на основе фундаментальной матрицы и требует большего числа отождествленных точек, он считается более приемлемым, поскольку реализуется с помощью линейного метода наименьших квадратов, причем число неизвестных оказывается фиксированным.

При этом, однако, в явном виде не восстанавливается взаимная ориентация камер и дальности до точек сцены. Если же на матрицу поворота  $\mathbf{R}$  и вектор переноса  $\mathbf{r}_o$  наложены дополнительные ограничения, то может быть удобнее использовать первый подход с явным восстановлением дальности (к примеру, именно это делается в задаче стереозрения).

#### *Фундаментальные матрицы при ограничениях, накладываемых на взаимное положение камер*

Рассмотрим случай переноса вдоль оси  $X$  без вращения. Этот случай похож на задачу стереозрения, но есть принципиальное отличие: мы не знаем модуля вектора смещения, то есть не знаем размера стереобазы. При

использовании фундаментальной матрицы в этом случае мы можем установить горизонтальность эпиполярных линий.

Действительно, при наличии только сдвига  $\Delta x$  имеем  $\Delta y = \Delta z = 0$  и  $r_{i,j} = 0$ , если  $i \neq j$  и  $r_{1,1} = r_{2,2} = r_{3,3} = 1$ . Несложно увидеть, что фундаментальная матрица должна принять вид:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\Delta x f_1' f_1' / f_2' \\ 0 & \Delta x f_1' & 0 \end{pmatrix}.$$

Рассмотрим, к примеру, изображения, представленные на рис. 3.12.

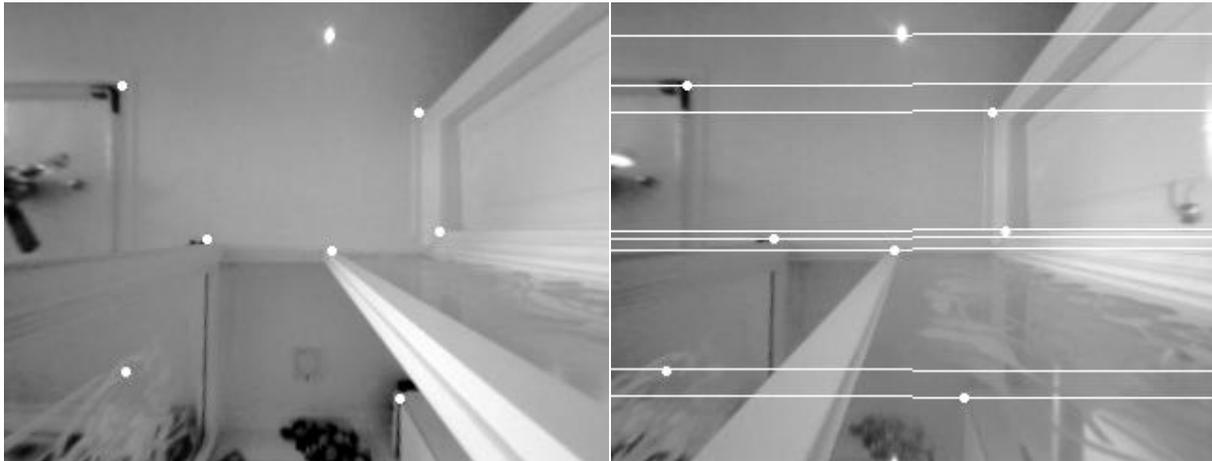


Рис. 3.12. Структура эпиполярных линий при перемещении камеры перпендикулярно оптической оси

Общий алгоритм оценивания фундаментальной матрицы для точек, представленных на рис. 3.12, может найти следующие значения, соответствующие по структуре перемещению вдоль оси  $X$ .

$$\begin{pmatrix} 6.8 \cdot 10^{-33} & -6.7 \cdot 10^{-18} & 3.4 \cdot 10^{-16} \\ 7.1 \cdot 10^{-18} & -2.0 \cdot 10^{-19} & 1.17 \cdot 10^{-2} \\ -9.8 \cdot 10^{-16} & -1.17 \cdot 10^{-2} & 2.3 \cdot 10^{-14} \end{pmatrix}.$$

Вращение вокруг оси  $Y$  также приводит к горизонтальному перемещению точек на изображении. Этот сдвиг, однако, одинаков для всех точек. Также одинаковым сдвиг будет при горизонтальном смещении камеры в случае равноудаленных объектов, то есть эти случаи неразличимы. Более того, если в дополнение к горизонтальному сдвигу присутствует вращение вокруг оси  $Y$ , то это никак не изменяет фундаментальной матрицы. Действительно, в этом случае мы получим

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\Delta x f_1' f_1' / f_2' \cos \alpha \\ 0 & \Delta x f_1' \cos \alpha & 0 \end{pmatrix},$$

где  $\alpha$  – угол вращения. А поскольку величина  $\Delta x$  находится с точностью до множителя, то этот множитель вберет в себя и коэффициент  $\cos \alpha$ .

При вращении вокруг оси  $Y$  имеем  $\Delta x = \Delta y = \Delta z = 0$ ,  $r_{2,2} = 1$ ,  $r_{1,1} = r_{3,3} = \cos \alpha$ ,  $r_{1,3} = \sin \alpha$ ,  $r_{3,1} = -\sin \alpha$ , что, вообще говоря, приводит к нулевой фундаментальной матрице. Таким образом, фундаментальная матрица в этом случае определена быть не может. Однако при наличии неточности в положении сопряженных точек (или при наличии других компонентов преобразования) и использовании метода наименьших квадратов некоторая фундаментальная матрица будет найдена, однако ее структура не будет нести информацию о том, какие вращения и сдвиги присутствуют.

В этом случае для любых сопряженных точек должно выполняться  $x'_2 = x'_1 + x_\alpha$ ,  $y'_1 = y'_2$ , для чего коэффициенты фундаментальной матрицы должны удовлетворять

$$(x'_1 + x_\alpha)(f_{11}x'_1 + f_{12}y'_1 + f_{13}) + y'_1(f_{21}x'_1 + f_{22}y'_1 + f_{23}) + (f_{31}x'_1 + f_{32}y'_1 + f_{33}) = 0.$$

при любых  $x'_1, y'_1$ . Перепишем это уравнение в виде

$$f_{11}x'_1x'_1 + (f_{12} + f_{21})x'_1y'_1 + f_{22}y'_1y'_1 + (f_{13} + f_{31} + x_\alpha f_{11})x'_1 + (f_{23} + f_{32} + x_\alpha f_{12})y'_1 + (f_{33} + x_\alpha f_{13}) = 0.$$

Чтобы это уравнение выполнялось для любых  $x'_1, y'_1$ , необходимо и достаточно соблюдение следующих равенств

$$f_{11} = f_{22} = 0, f_{12} = -f_{21}, f_{13} = -f_{31}, f_{23} = -f_{32} - x_\alpha f_{12}, f_{33} = -x_\alpha f_{13}.$$

Таким образом, параметры  $f_{21}, f_{31}, f_{32}$  оказываются абсолютно произвольными и могут принимать любые значения в зависимости от случайных факторов.

К примеру, если задать некоторое количество (больше восьми) случайных точек  $x'_{1,i}, y'_{1,i}$  и задать сопряженные им точки как  $x'_{2,i} = x'_{1,i} - 20$ ,  $y'_{2,i} = y'_{1,i}$  с добавлением случайных ошибок в пределах одного пикселя, то методом наименьших квадратов может быть восстановлена, например, матрица

$$\begin{pmatrix} -7.1 \cdot 10^{-7} & -3.6 \cdot 10^{-4} & 4.9 \cdot 10^{-2} \\ 3.6 \cdot 10^{-4} & -1.6 \cdot 10^{-6} & -9.7 \cdot 10^{-2} \\ -4.9 \cdot 10^{-2} & 9.0 \cdot 10^{-2} & 1.0 \end{pmatrix}.$$

Видно, что  $f_{11} \approx f_{22} \approx 0$ ,  $f_{13} \approx f_{33}/20$ ,  $f_{12} \approx -f_{21}$ ,  $f_{13} \approx -f_{31}$ ,  $f_{23} \approx -f_{32} + 20f_{12}$ .

Хотя эти параметры и задают эпиполярные линии, проходящие через сопряженные точки, они не содержат информацию о том, что положения камер могут быть связаны вращением вокруг оси  $Y$  на определенный угол.

Интересен также случай переноса вдоль оптической оси без вращения (матрица вращения является единичной). В этом случае фундаментальная матрица примет вид

$$\begin{pmatrix} 0 & -\Delta z f'_1/f'_2 & 0 \\ \Delta z f'_1/f'_2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Эпиполярные линии будут выражаться в форме  $x'_2 / y'_2 = x'_1 / y'_1$ .

Матрица может быть определена с точностью до постоянного множителя, поэтому  $\Delta z$  здесь также будет неизвестно.

Посмотрим, какого вида фундаментальную матрицу можно восстановить, если имеющиеся сопряженные точки лежат на эпиполярных линиях вида  $x'_2 / y'_2 = x'_1 / y'_1$ . Это означает, что  $x'_2 = ax'_1, y'_2 = ay'_1$ . Тогда имеем

$$ax'_1(f_{11}x'_1 + f_{12}y'_1 + f_{13}) + ay'_1(f_{21}x'_1 + f_{22}y'_1 + f_{23}) + (f_{31}x'_1 + f_{32}y'_1 + f_{33}) = 0.$$

Откуда

$$\begin{aligned} f_{11}ax'_1x'_1 + a(f_{12} + f_{21})x'_1y'_1 + f_{22}ay'_1y'_1 + \\ (f_{13}a + f_{31})x'_1 + (f_{23}a + f_{32})y'_1 + f_{33} = 0. \end{aligned}$$

Это уравнение должно выполняться для любых значений  $x'_1, y'_1$ , поэтому необходимо, чтобы было верно

$$f_{11} = f_{22} = f_{33} = 0, f_{12} = -f_{21}, af_{13} = -f_{31}, af_{23} = -f_{32}.$$

Кроме того, величина  $a$  может меняться (за исключением вырожденных случаев, когда все точки расположены на одной глубине). Выполнение этих уравнений при любых значениях  $a$  возможно, только если  $f_{13} = f_{31} = f_{23} = f_{32} = 0$ . Иными словами, при движении вдоль оптической оси и невырожденном расположении сопряженных точек будет восстановлена матрица, в которой  $f_{12} = -f_{21}$ , а остальные коэффициенты равны нулю.

Несложно заметить, что, если нет сдвигов ( $\Delta x = \Delta y = \Delta z = 0$ ), то не только при вращении вокруг оси  $Y$ , но и при произвольном вращении фундаментальная матрица оказывается нулевой. Это неудивительно, поскольку при вращении видимые объекты смещаются в плоскости изображения вне зависимости от дальности до них, иными словами, эпиполярная линия вырождается в точку, что может быть описано только нулевой фундаментальной матрицей. Это ограничивает возможность использования подхода к сопоставлению изображений трехмерных сцен с использованием фундаментальной матрицы в случае, когда меняется только направление наблюдения, но не меняется точка наблюдения.

Как и при восстановлении фундаментальной матрицы при вращении вокруг оси  $Y$ , при восстановлении фундаментальной матрицы в случае другого вида вращения нулевая фундаментальная матрица построена не будет (за исключением случаев абсолютно точного задания координат сопряженных точек). Вместо этого будет восстановлена некоторая (произвольная) матрица, задающая эпиполярные линии, проходящие через сопряженные точки. Произвольность восстановления этой матрицы продемонстрирована на рис. 3.13.

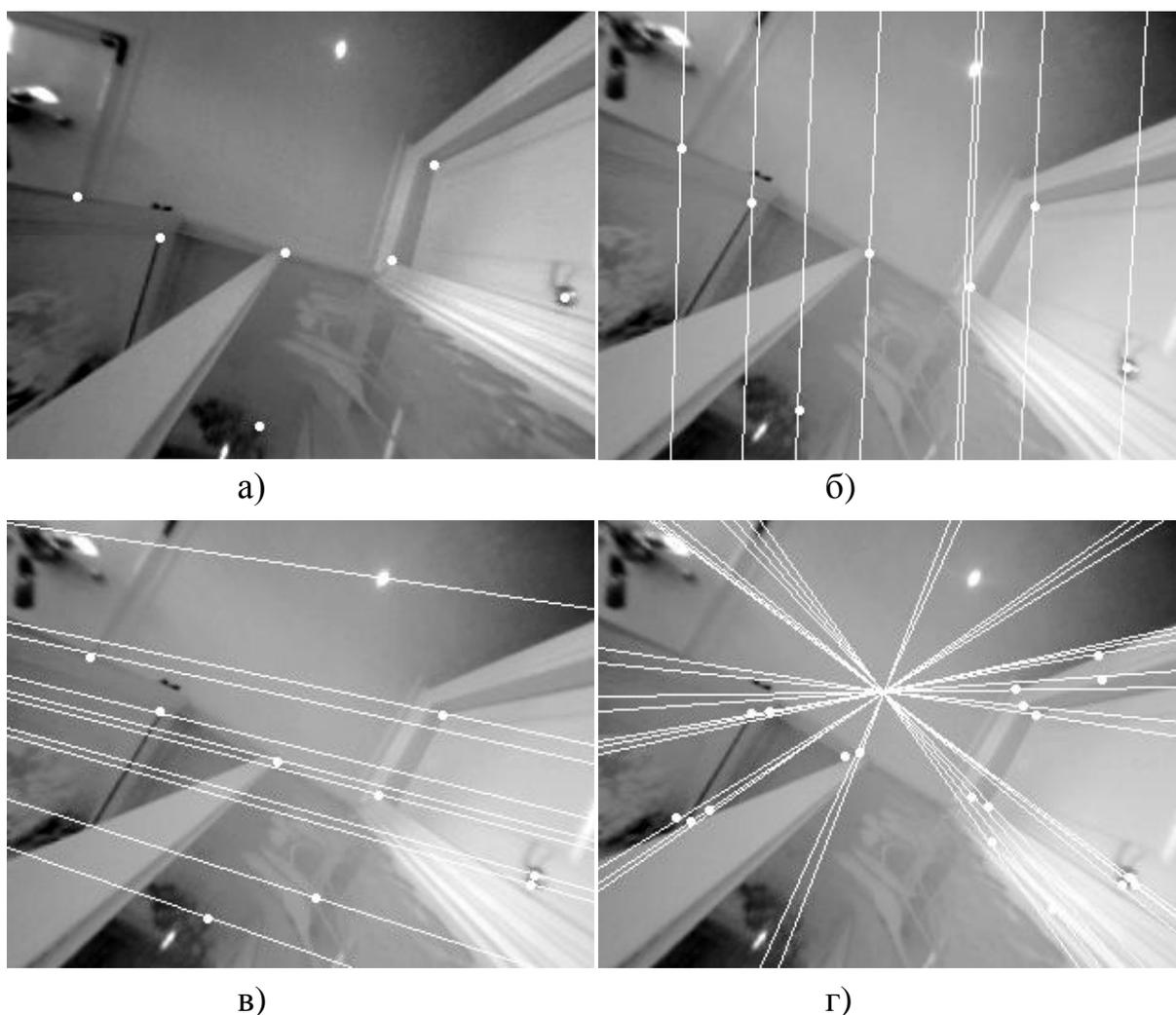


Рис. 3.13. Пример построения различных систем эпиллярных линий при определении фундаментальной матрицы по разным наборам сопряженных точек при одном и том же пространственном преобразовании (вращении)

На этом рисунке представлены примеры восстановления фундаментальной матрицы с помощью метода наименьших квадратов по разным наборам точек. На рис. 3.13а и 3.13б представлен вариант системы эпиллярных линий, найденных по 8 точкам (изображения получены при вращении камеры вокруг оси  $Z$ ). На рис. 3.13в показано изменение результата при добавлении двух дополнительных точек. На рис. 3.13г система эпиллярных линий получена по большему числу точек. Необходимо подчеркнуть, что все эти эпиллярные линии не противоречат сопряженным точкам, что говорит о неоднозначности решения проблемы восстановления фундаментальной матрицы в отсутствие сдвигов.

Отметим, что вращение вокруг оси  $Z$  в варианте 3.13г дает те же эпиллярные линии, что и то же вращение, но с дополнительным смещением по оси  $Z$ . Таким образом, если учитывать только попадание точек на эпиллярные линии, но не учитывать их положение на этих линиях, то эти два случая различить невозможно.

Поскольку без знания расстояния до объектов может быть восстановлено только направление вектора переноса, но не его модуль, то его

значение может быть приравнено к единице. При этом случай чистого вращения будет эквивалентен случаю бесконечной удаленности до объектов (а поскольку дальности не восстанавливаются, эти случаи будут в смысле фундаментальной матрицы неразличимы).

Таким образом, если заранее известны ограничения на взаимное расположение камер, то вместо подхода на основе фундаментальной матрицы имеет смысл использовать специальные подходы, учитывающие в явном виде имеющиеся ограничения. Кроме того, подход на основе фундаментальной матрицы удобен для сопоставления (распознавания) изображений трехмерных сцен, но не позволяет получить в явном виде относительную ориентацию камер и расстояния до точек сцены.

#### *Структурный подход к отождествлению сопряженных точек*

Классическим способом определения взаимного положения и ориентации камер является использование специальных фотограмметрических меток (рис. 3.14).

Как правило, фотограмметрические метки легко выделяются на изображении и могут быть без особых трудностей отождествлены друг с другом, что позволяет определить относительную ориентацию камер. Кроме того, известные реальные расстояния между метками помогают оценить и истинное расстояние между камерами, что затем позволяет решать задачу стереозрения с определением абсолютной дальности до точек сцены.



Рис. 3.14. Пример производственной сцены с системой фотограмметрических меток на переднем плане для решения проблемы

Помимо этого, определение ориентаций камеры относительно системы меток позволяет решить проблему *внешней ориентации*: далее системой компьютерного зрения координаты всех объектов сцены могут пересчитываться в координаты относительно системы фотограмметрических меток. Это может использоваться, например, для управления манипулятором

робота, положение которого известно в той же внешней системе координат (таким образом, может быть выполнено преобразование «глаз-рука»).

Однако использование фотограмметрических меток для решения проблемы ориентации камер удобно в основном только в промышленных условиях, но и там далеко не всегда является подходящим решением. Во многих прочих случаях (например, для бытовых роботов) использование меток вообще неприемлемо.

Таким образом, необходимо иметь возможность отождествления сопряженных точек на изображениях трехмерных сцен. Вполне естественным здесь выглядит структурный подход.

В случае структурного сопоставления изображений плоских сцен производился перебор возможных соответствий, для чего требовалось порядка  $K^{2+M/2}$  операций, где  $K$  – число элементов на одном изображении,  $M$  – число параметров пространственного преобразования. Значение  $M/2$  в показателе возникало потому, что одно соответствие давало два уравнения. Для аффинного преобразования число операций получалось пропорциональным  $K^5$ . Здесь же одно соответствие дает одно уравнение, поэтому при семи неизвестных параметрах число операций будет порядка  $K^9$ , что очень много. Перебор вариантов соответствия в случае трехмерных сцен вряд ли возможен.

Одна из возможностей заключается в выделении на изображении отдельных поверхностей (рис. 3.15) и разделении всех структурных элементов на группы по их принадлежности этим поверхностям. При условии, что поверхности плоские и находятся не слишком близко к камерам, сопряженные поверхности будут связаны аффинным преобразованием (однако для разных пар поверхностей параметры этого преобразования будут различными). Поскольку заранее неизвестно, какие поверхности являются сопряженными, оказывается необходимым также осуществлять перебор всех пар групп структурных элементов. Таким образом, процедура сопоставления выполняется иерархически: на нижнем уровне сопоставляются структурные элементы внутри групп, а на верхнем уровне сопоставляются группы.

Реализация указанного подхода является, однако, весьма трудоемкой, а процесс сопоставления оказывается весьма ресурсоемким. Кроме того, в рамках данного подхода необходимо разделять изображение на плоские поверхности. Хорошего разделения в общем случае далеко не всегда удастся добиться существующими методами сегментации. В то же время данный подход может оказаться наиболее робастным (при условии хорошей сегментации), когда изображения получены при различных условиях освещения.



Рис. 3.15. Пример разделения изображения трехмерной сцены на плоские поверхности

При обсуждении структурного сопоставления упоминалась также возможность использования дополнительных признаков для сокращения числа возможных соответствий между структурными элементами. Проблема заключается в том, чтобы построить признаки, инвариантные не только к геометрическим, но и яркостным искажениям. При рассмотрении структурного сопоставления мы не рассматривали эту проблему, поскольку она не всегда разрешима из-за существенной изменчивости изображений, и попытка использования локальных признаков привела бы к построению неуниверсальных методов сопоставления. При сопоставлении трехмерных сцен привлечение локальных признаков оказывается чуть ли не единственным способом решения проблемы сопоставления, хотя при этом и приходится отказываться от возможности работы методов сопоставления в условиях существенной изменчивости.

#### *Системы инвариантных признаков*

Рассмотрим некоторые подходы к построению систем инвариантных признаков. Процедуры, используемые для генерации множества признаков изображения, как правило, состоят из следующих основных шагов.

1. Обнаружение пространственно-масштабных экстремумов:

- а) строится пирамида разрешений путем сглаживания изображения с гауссианами с разной дисперсией или с прямоугольными окнами с разной шириной;

- б) производится дифференцирование изображения на каждом уровне разрешения, например, путем вычисления функции DoG (Difference of Gaussian, «разности гауссиан») для идентификации потенциальных точек интереса, инвариантных к масштабу и ориентации;
  - в) на построенном пространственно-масштабном поле производится поиск локальных максимумов – точек интереса.
2. Локализация ключевых точек:
- а) для каждой точки интереса производится интерполяция для определения положения и масштаба;
  - б) осуществляется выбор ключевых точек из всего множества точек интереса на основе оценок их стабильности.
3. Приписывание ориентации:
- а) одна или более ориентаций приписываются каждой ключевой точке на основе локальных направлений градиента изображения. Все последующие операции выполняются на изображении, которое преобразуется по отношению к приписанным ориентации, масштабу и положению каждой ключевой точки, что влечет инвариантность к этим преобразованиям.
4. Описание ключевых точек:
- а) локальные градиенты изображения или какие-либо другие характеристики измеряются на выбранном масштабе в области каждой ключевой точки;
  - б) исходные измеренные значения трансформируются в представление, которое допускает значительный уровень локального искажения формы и изменения освещения.

Рассмотрим чуть более детально метод SIFT (Scale Invariant Feature Transform), имеющий такое название, поскольку в нем изображение преобразуется в масштабно-инвариантную систему координат относительно локальных признаков. В этом методе пространственно-масштабная функция от изображения  $f(x,y)$  определяется как свертка

$$L(x, y, \sigma) = f(x, y) \otimes G(x, y, \sigma),$$

где  $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$ .

Масштабный множитель  $\sigma$  принимает значения в некотором диапазоне. Эти значения вычисляются путем последовательного умножения на некоторую постоянную величину  $k$ . Два соседних масштабных уровня получаются при значениях масштабного множителя, равных  $\sigma$  и  $k\sigma$ .

Функция DoG от изображения определяется как разность

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \otimes f(x, y).$$

Эта функция аппроксимирует оператор Лапласа от изображения, сглаженного с гауссианами разной ширины, поскольку

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G.$$

Далее в пространстве  $(x, y, \sigma)$  производится поиск локальных максимумов и минимумов функции  $D$ . Как правило, таких локальных максимумов крайне много, поэтому из них выбираются наиболее стабильные, что делается следующим образом.

Сначала уточняется положение максимума путем интерполяции функции  $D(x, y, \sigma)$  с помощью разложения ее в ряд Тейлора:

$$D(\mathbf{r}) = D + \frac{\partial D}{\partial \mathbf{r}} \mathbf{r} + \frac{1}{2} \mathbf{r}^T \frac{\partial^2 D}{\partial \mathbf{r}^2} \mathbf{r},$$

где  $\mathbf{r} = (x, y, \sigma)$ .

Положение экстремума может быть определено путем приравнивания нулю производной от  $D$ . Можно получить

$$\mathbf{r}^* = \left[ \frac{\partial^2 D}{\partial \mathbf{r}^2} \right]^{-1} \frac{\partial D}{\partial \mathbf{r}}.$$

Частные производные вычисляются по значению функции в соседних точках. Значение функции в экстремуме  $D(\mathbf{r}^*)$  используется для определения, какие из максимумов являются стабильными.

В данном подходе (SIFT) также оказывается необходимым устранять края, так как на всем их протяжении функция  $D$  имеет большие значения.

Далее определяется характерная ориентация в каждой ключевой точке, что можно сделать различными способами. Простейший способ заключается в оценке направления градиента в данной точке:

$$\theta(x, y, \sigma) = \tan^{-1} \left[ \frac{\partial L(x, y, \sigma)}{\partial y} / \frac{\partial L(x, y, \sigma)}{\partial x} \right].$$

Может использоваться усреднение направлений ориентации градиента (взвешенное по модулю градиента) в окрестности текущей точки.

В действительности, положение точек интереса с характерными значениями масштаба и ориентации могут быть найдены и другими способами. Основной же вопрос заключается в построении системы инвариантных признаков, с помощью которых можно было бы идентифицировать каждую точку.

Набор параметров: положение, масштаб и ориентация, – могут быть использованы для вращения и масштабирования окрестности каждой ключевой точки изображения, что даст фрагмент изображения, инвариантный (при условии инвариантности определения этих параметров) к соответствующим геометрическим искажениям. В определенной степени это означает и инвариантность по отношению к смене ракурса.

Очевидным подходом для отождествления ключевых точек является дальнейшее вычисление корреляции между фрагментами, взятыми вокруг соответствующих точек. Однако эта корреляция очень чувствительная к некомпенсированным компонентам аффинного преобразования и погрешностям в локализации точек.

В SIFT предлагается следующий подход к построению дескрипторов ключевых точек, малочувствительных к подобным некомпенсированным искажениям. На выбранном масштабном уровне берутся значения модуля и направления градиентного поля в окне вокруг ключевой точки. При этом направления градиента поворачиваются на характерный угол, найденный для текущей ключевой точки. Каждому значению градиента приписывается вес, убывающий по нормальному закону с дисперсией, равной половине ширины окна.

Окно вокруг точки интереса разбивается на области  $4 \times 4$  пикселя, в каждой из которых строится гистограмма градиентов, содержащая 8 бинов. В эту гистограмму для каждого направления записывается средняя величина модуля градиенты с данным направлением. Таких областей берется также  $4 \times 4$ . Таким образом, рассматривается окно  $16 \times 16$  пикселей, по которому строится  $4 \times 4 \times 8 = 128$  признаков. Для устранения влияния изменения освещения вектор признаков нормируется. Число бинов в гистограмме и размеры окон являются настроечными параметрами метода и могут быть изменены.

В альтернативном подходе, называемом SURF (Speeded Up Robust Features), вместо сглаживания с гауссианой использует сглаживание с прямоугольным окном, что позволяет заметно увеличить быстродействие процедур выделения максимумов в пространстве-масштабе. Дескриптор в данном подходе также вычисляется по-другому, чем в SIFT. Здесь локальное окно (после компенсации масштаба и поворота) разбивается на  $4 \times 4$  области, состоящие из  $5 \times 5$  пикселей. Для каждой точки окна вычисляются отклики на вейвлеты Хаара (их  $x$ - и  $y$ -составляющие), после чего полученные отклики усредняются по  $5 \times 5$  областям. Усреднению подвергаются как сами значения откликов, так и их модули. В результате, каждая область описывается 4 признаками, что дает  $4 \times 4 \times 4 = 64$  признака в дескрипторе ключевой точки.

На основе SIFT или SURF дескрипторов ключевые точки могут непосредственно отождествляться методом ближайшего соседа, в котором для каждой точки первого изображения находится ближайшая (в признаковом пространстве) точка второго изображения. В простых случаях (при малых яркостных и ракурсных искажениях) этого может быть вполне достаточно. Однако такое парное отождествление точек приводит к достаточно большому числу неверных отождествлений, в связи с чем оказывается необходимым учитывать взаимное расположение ключевых точек в процессе их отождествления, то есть использовать структурное сопоставление.

Существует большое разнообразие вариантов реализации структурного сопоставления ключевых точек, обладающих инвариантными дескрипторами. Во-первых, можно отобрать небольшое число (8-12) наиболее надежных отождествлений и построить на их основе начальную оценку фундаментальной матрицы, которую далее использовать в процессе отождествления. Этот прием позволяет без существенных вычислительных затрат получить согласованные отождествления сопряженных ключевых

точек, но ошибка в начальной гипотезе приведет к неправильному сопоставлению.

Во-вторых, можно использовать преобразование  $H_0$ : каждая сопоставленная пара ключевых точек задает сдвиг, взаимные поворот и масштаб изображений, поскольку к каждой точке приписываются пространственные координаты, характерные масштаб и ориентация. С помощью преобразования  $H_0$  можно искать параметры глобального аффинного преобразования, которое, хотя и не описывает в точности взаимное пространственное преобразование двух изображений, но в общих чертах задает перемещение всех точек сцены. Найденное аффинное преобразование может использоваться для поиска взаимно согласованных отождествлений ключевых точек.

В-третьих, может применяться структурное сопоставление, выполняемое с помощью поиска по дереву вариантов соответствий ключевых точек. При этом большое число ветвей на дереве вариантов может быть отсечено на основе результатов сравнения дескрипторов ключевых точек.

### **Вопросы и упражнения**

1. Какие величины связывает уравнение относительной ориентации?
2. Сколько уравнений и неизвестных добавляет одна пара сопряженных точек в задаче относительной ориентации?
3. Какой вид эпиполярные линии имеют в случае движения камеры вдоль оптической оси, поперек нее, при вращении камеры, в случае двух произвольных проекций?
4. Что такое фундаментальная матрица? Чем отличаются задачи определения фундаментальной матрицы и относительной ориентации?
5. Когда фундаментальная матрица оказывается неопределенной?
6. В чем заключается задача относительной ориентации?
7. Для чего используются системы инвариантных признаков?
8. Каковы общие идеи подходов SIFT и SURF?

### **3.3. Методы восстановления трехмерной формы и дальности**

Одной из базовых задач компьютерного зрения является восстановление трехмерной информации об объектах по плоским изображениям, что имеет очень широкое практическое применение в робототехнике, автономной навигации, неразрушающем контроле качества продукции и т.д.

Как правило, в задачах компьютерного зрения изображения формируются на основе света, отраженного от набора непрозрачных поверхностей (случаи полупрозрачных сред, таких как стёкла, рассматриваются весьма редко), то есть каждая точка изображения соответствует фиксированной точке сцены. Это отличается от задач

томографии, в которых также восстанавливается трехмерная информация, но электромагнитное излучение проходит через слой вещества, взаимодействуя с ним.

Методы восстановления трехмерной информации в компьютерном зрении весьма разнообразны. Ниже приведены некоторые наиболее типичные из них:

- активные методы
  - структурированная подсветка,
  - дальность из фазы,
- пассивные методы
  - стереозрение,
  - дальность из двух произвольных проекций,
  - форма из движения,
  - форма из текстуры,
  - форма из затенения,
  - 3D геометрия из структуры,
  - восстановление дальности из фокусировки и дефокусировки.

Из перечисленных методов выше уже были рассмотрены методы стереозрения и восстановления дальности из двух произвольных проекций, что относится к пассивным методам.

Активные методы, как правило, используют более громоздкую и дорогую аппаратуру, но требуют более простых алгоритмов анализа изображений для получения информации о дальности (и часто результат их работы оказывается надежнее).

#### *Метод структурированной подсветки*

Одним из классических активных методов является метод структурированной подсветки, при котором на сцену проецируется некоторый регулярный световой паттерн (см. рис. 3.16), по искажению которого можно судить о трехмерной форме объектов и дальности до них.

Сейчас часто привлекается структурированная подсветка сцены в диапазоне длин волн, отстоящем от основного диапазона естественного освещения сцены (например, инфракрасная подсветка), что облегчает идентификацию точек. Кроме того, может использоваться цветная подсветка, позволяющая выполнять измерения одновременно в нескольких каналах (в каждом цветовом канале проецируется свой паттерн). В целом, данный подход позволяет получить достаточно точную оценку формы объектов (см. рис. 3.18).

Метод структурированной подсветки имеет определенные ограничения и особенности. Структурированность подсветки означает, что не для всех точек сцены может быть произведено измерение дальности при проецировании единичного паттерна. В связи с этим оказывается необходимым последовательно проецировать несколько различных паттернов (например, полос разной толщины), что замедляет работу метода, но может дать точную оценку формы.

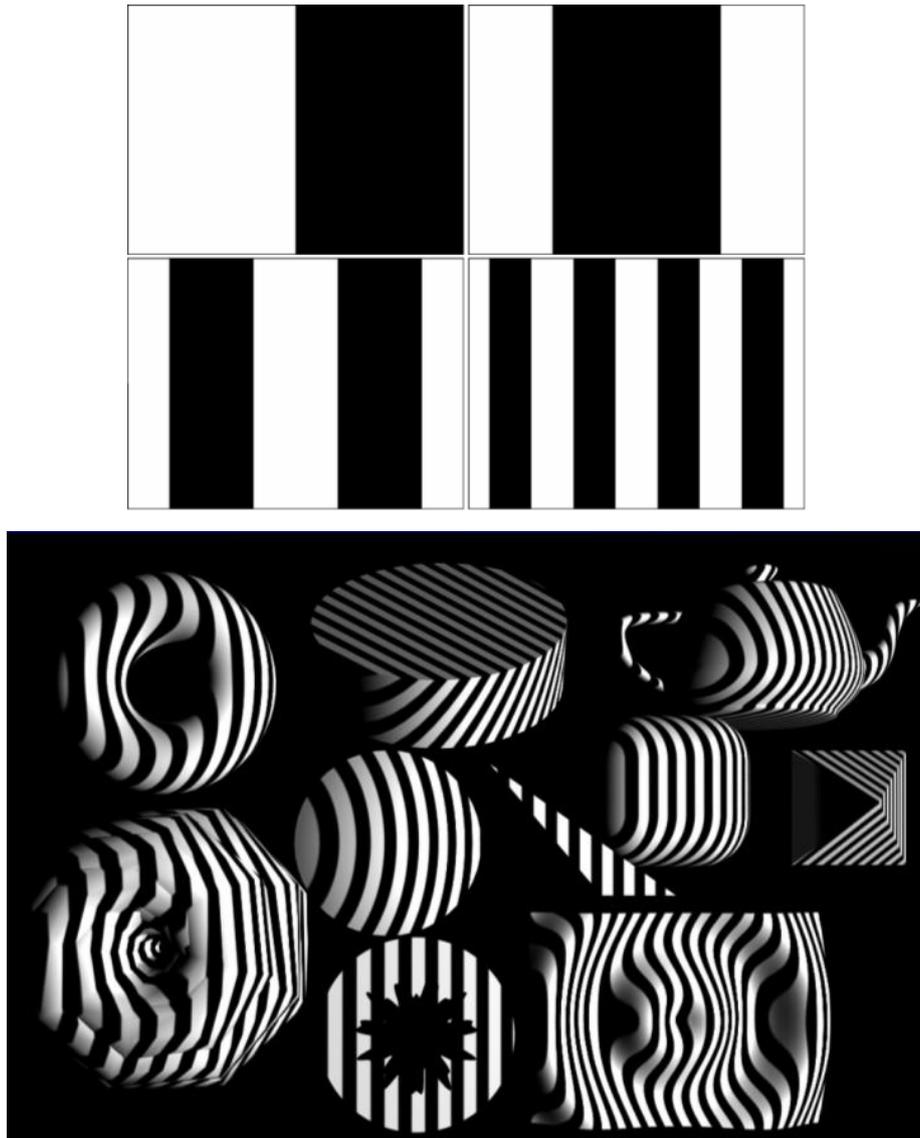


Рис. 3.16. Примеры проецируемых паттернов и результатов освещения

Способ определения дальности здесь аналогичен стереозрению и заключается в установлении направлений на одну и ту же точку с двух ракурсов (см. рис. 3.17). Однако вместо второй камеры используется источник освещения. Структурирование этого освещения позволяет сопоставить точку на изображении, формируемом камерой, с одним из лучей, формируемых источником освещения.

В качестве дополнительного ограничения выступает чувствительность метода к дальности до объектов сцены. Как правило, объекты должны быть на вполне определенном расстоянии. Если же объект окажется слишком близко, то вся структурированная подсветка будет проецироваться лишь на его небольшой участок. Напротив, если объект окажется слишком далеко, то один элемент структуры подсветки (например, полоса) может прийтись на весь объект. В связи с этим, метод структурированной подсветки имеет ограниченное применение. Наиболее частым является его применение при 3D-сканировании объектов или на ленте конвейера, где система технического зрения располагается на фиксированном удалении от измеряемых объектов.

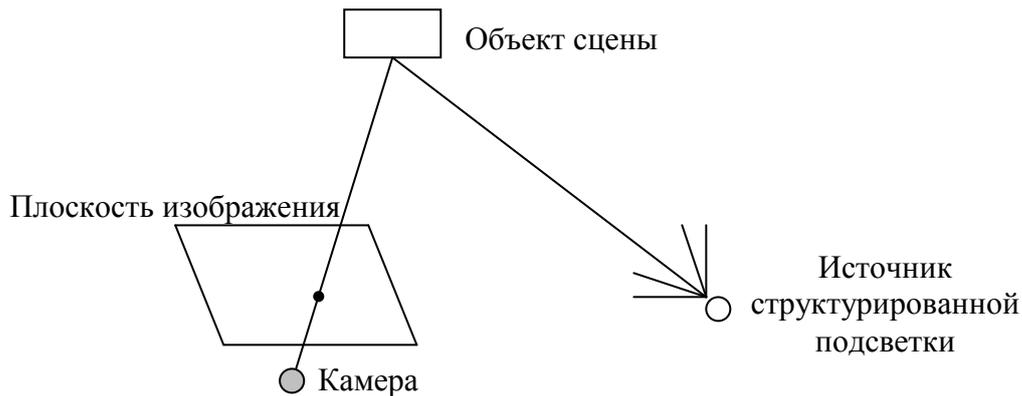


Рис. 3.17. Определение дальности до объекта методом структурированной подсветки путем триангуляции по двум направлениям

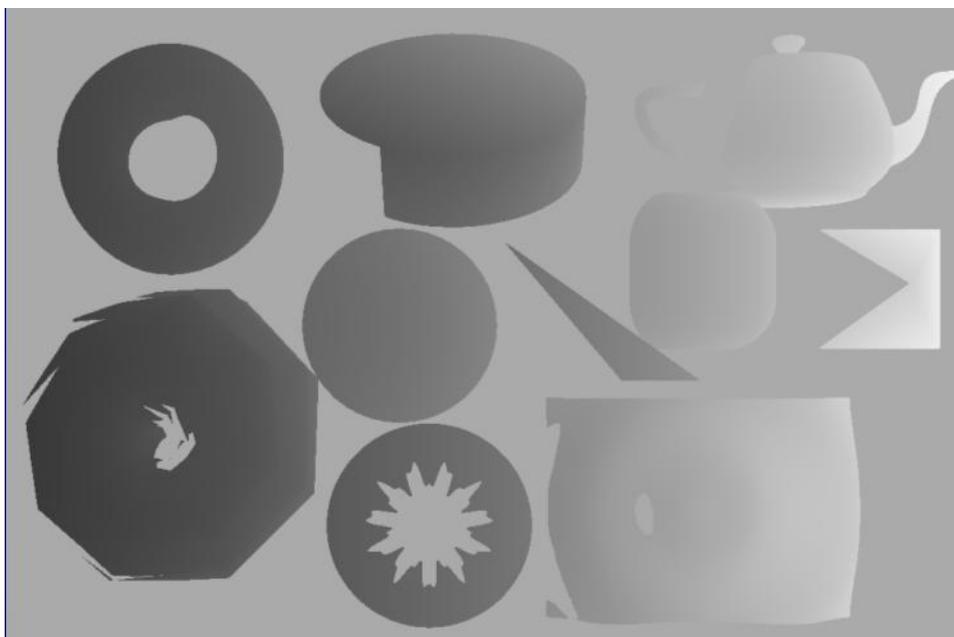


Рис. 3.18. Пример результатов измерения формы объектов методом структурированной подсветки

### *Дальность из фазы*

Другим классическим активным методом определения дальности является измерение разницы во времени между моментом генерации сигнала и моментом его регистрации после отражения. Тогда расстояние до точки отражения будет равно  $v_s(t_r - t_g)/2$ , где  $v_s$  – скорость распространения сигнала,  $t_g$  – момент времени генерации сигнала,  $t_r$  – момент времени регистрации отраженного сигнала.

В случае малой скорости распространения сигнала (например, в эхолокаторах) временная разница может быть измерена непосредственно. Однако при использовании электромагнитного излучения зачастую это сделать проблематично, в связи с чем используется измерение фазы волны

(для чего необходимо применять когерентное излучение). Измерение фазы волны часто выполняется интерферометрически.

Лазерные дальномеры и радары с синтезированной апертурой достаточно широко используются как источник трехмерной информации в системах компьютерного зрения. Однако процедуры восстановления дальности с использованием когерентного излучения часто рассматриваются вне проблематики компьютерного зрения, и здесь мы их подробно разбирать не будем.

Пассивные методы восстановления трехмерной информации можно разделить на две группы: методы, использующие несколько изображений, полученных с разных ракурсов, и методы, использующие одной изображение. К первой группе относятся методы стереозрения, дальности из нескольких проекций (движения наблюдателя), формы из движения объекта. Ко второй группе относятся методы восстановления формы из текстуры, затенения и структуры. Есть также ряд специфических методов, использующих несколько изображений, снятых с одного ракурса, но при разных условиях. Это методы глубины из фокусировки и дефокусировки, а также метод фотометрического стерео. Последний метод тесно связано с восстановлением формы из затенения, и будет рассмотрен в следующем разделе. Методы, использующие анализ движения, имеют много общего с уже рассмотренным методом восстановления дальности из нескольких проекций. Проблема анализа движения также будет затронута в следующем разделе. Здесь же рассмотрим некоторые специфические методы восстановления трехмерной информации.

#### *Форма из текстуры*

Одним из подходов к восстановлению трехмерной формы объектов по одному плоскому снимку, является подход «форма из текстуры». Этот подход использует повторяемость элементов текстуры, текселов, видимые характеристики которых меняются в зависимости от локальной ориентации поверхности и расстояния до нее. В некотором смысле это соответствует съемке одного и того же тексела с разных ракурсов, однако из-за малости размеров текселов и их изменчивости обычные методы сопоставления трехмерных сцен здесь не применимы (рис. 3.19).

С одной стороны, текстурная информация не несет точной информации о форме объектов (в особенности это относится к естественным текстурам, обладающим большой вариативностью текселов) и является сложной в использовании из-за большого разнообразия текстур. С другой стороны, часто текстура оказывается чуть ли не единственным источником трехмерной информации. В частности, человеком этот источник информации используется весьма активно при восприятии как естественных, так и искусственных изображений (см. рис. 3.19).

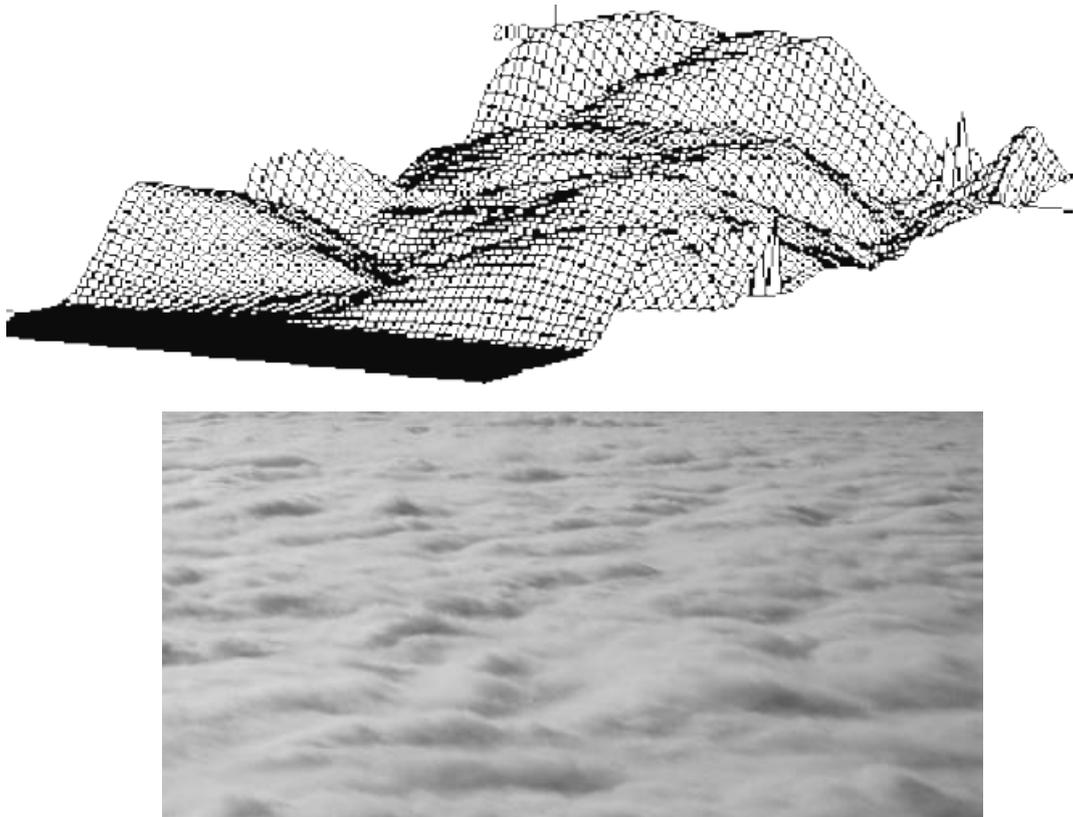


Рис. 3.19. Примеры искусственных и естественных текстурированных изображений, по которым человеческая зрительная система способна восстановить форму и ориентацию поверхности

Представим, что текстура состоит из идентичных текселов. Локальное изменение ориентации поверхности приведет к сжатию тексела в направлении наклона пропорционально косинусу угла. Таким образом, если можно определить локальную ориентацию текселов и коэффициент их сжатия, то можно определить и локальную ориентацию поверхности. На рис. 3.20 приведены примеры искажения прямоугольных текселов: изображение тексела получается в результате применения к нему аффинного преобразования, определяемого ориентацией поверхности. Оценка параметров аффинного преобразования, переводящего два соседних тексела друг в друга, позволяет оценить локальное изменение ориентации поверхности. Однако в случае симметричных текселов оценка параметров преобразования будет неоднозначной.

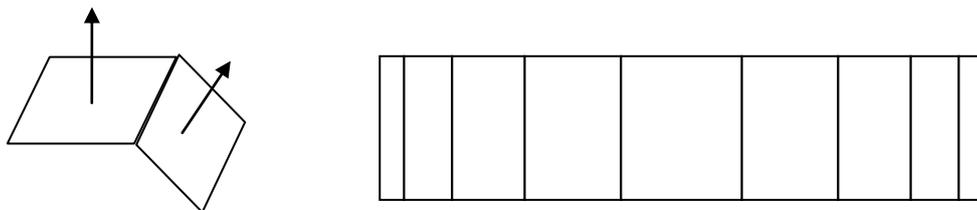


Рис. 3.20. Пример локального искажения текселов при изменении ориентации поверхности

Кроме того, если точки поверхности расположены на существенно разном удалении от наблюдателя, то будут также меняться и размеры

текселов обратнопропорционально расстоянию до соответствующих точек, что также позволяет найти наклон поверхности. Стоит, однако, иметь в виду, что наклон плоской поверхности, расположенной на значительном удалении от наблюдателя, найти на основе текстурного анализа будет невозможно, если заранее неизвестна ненаклонная форма текселов, поскольку на такой поверхности все текселы будут обладать одинаковым искажением, вызванным постоянным наклоном.

Методы, применяемые для оценивания искажений текселов, во многом похожи на методы построения системы локальных признаков, инвариантных к аффинным искажениям, поскольку в этих методах сначала как раз и производится оценка характерных изотропного и анизотропного масштабов, ориентации и скоса. Это может выполняться путем вейвлет-преобразования с определением значений параметров, при которых отклик максимален.

Более простой задачей, чем полное восстановление формы, является задача определения наличия резкого изменения ориентации поверхности объекта (изгибов или перекручивания). К примеру, на рис. 3.21 представлена выборка гибких объектов разной формы с текстурированной поверхностью.

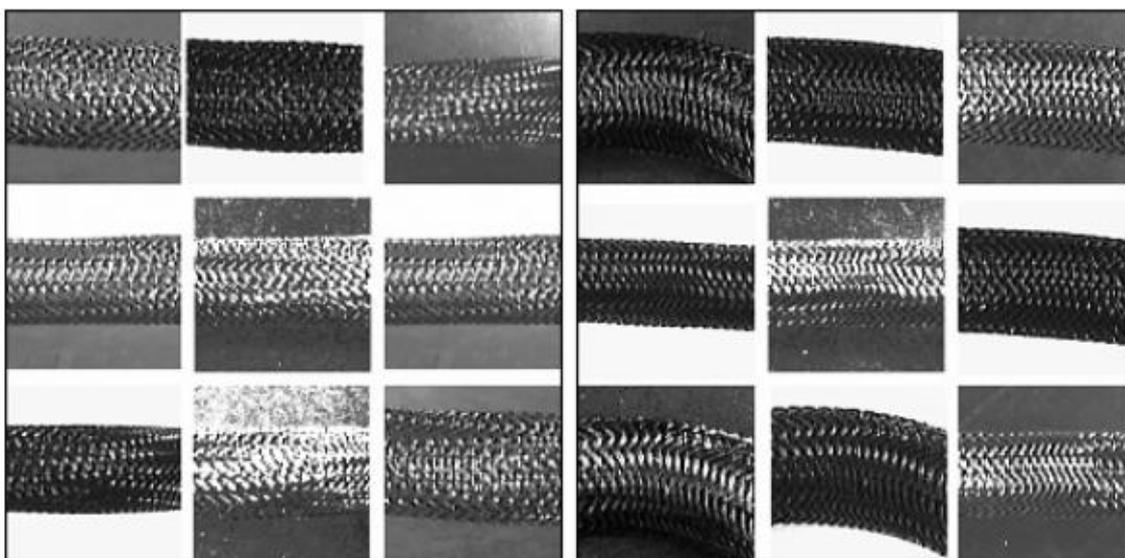


Рис. 3.21. Примеры изображений шланга с разной кривизной (в ряде случаев направление изгиба совпадает с направлением к наблюдателю)

Задача определения наличия изгибов или перекручивания может решаться методами распознавания образов на основе глобальных спектральных признаков. На основе обучающей выборки формируются классы образов, соответствующие случаям отсутствия или наличия искривлений поверхности. При выборе подходящих признаков классы оказываются разделимыми.

#### *Дальность из фокусировки и дефокусировки*

Одной из особенностей реальных объективов является конечная глубина резкости. Как результат, часть сцены оказывается «не в фокусе». Результат действия расфокусировки может быть описан через свертку исходного изображения с некоторым ядром, размеры которого соответствуют степени

расфокусировки (отклонения точки фокусировки от положения объекта). Ядро свертки может оказаться достаточно сложного вида (многие объективы имеют свой характерный «рисунок»), однако в нулевом приближении оно может быть описано гауссианой с некоторой дисперсией.

Простейший способ определения дальности из фокусировки заключается в постепенном изменении фокусного расстояния (с которым изменяется дальность фокусировки в предметном пространстве) с установлением значения, дающего максимальную резкость изображения. Критерий резкости естественным образом проистекает из того факта, что при свертке с гауссианой (и многими другими ядрами, которые могут возникать вследствие расфокусировки) происходит снижение резкости.

Количественно резкость изображения может быть оценена путем анализа распределения энергии в его пространственном спектре. Простейший коэффициент резкости может быть введен как

$$C_{sh}^2 = \frac{\sum_{u=0}^{N-1} \sum_{v=0}^{N-1} [u^2 + v^2] |\tilde{f}(u, v)|^2}{\sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |\tilde{f}(u, v)|^2},$$

где  $\tilde{f}(u, v)$  – пространственный спектр изображения  $f$  размера  $N$ . Величина  $C_{sh}$  будет показывать характерную частоту изображения.

Следует, однако, иметь в виду, что данная величина позволяет различить резкость одного и того же изображения, но не позволяет отличить резкость двух разных изображений (например, для пар фрагментов на рис. 3.22а,б и 3.22в,г эта величина позволяет выбрать фрагмент большей резкости, а пар фрагментов на рис. 3.22а,г и 3.22б,в – нет). Кроме того, нужно иметь в виду, что при вычислении спектра изображения необходимо устранять эффект края, а также симметрию спектра.

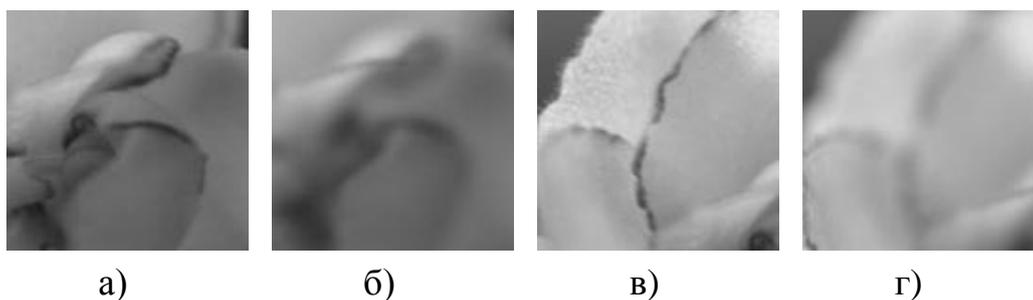


Рис. 3.22. Примеры фрагментов изображения с разной резкостью

Конкретные значения этого коэффициента не несут особой смысловой нагрузки, однако путем их сравнения для одного и того же изображения, но разных фокусных расстояний можно установить момент наибольшей резкости, что является достаточным для определения дальности до сцены и может использоваться также для автоматической фокусировки цифровых камер.

В случае, когда глубина резкости (определяемая также относительным отверстием) оказывается меньше глубины сцены, разные точки сцены будут иметь разную резкость. Это дает инструмент для построения карты дальности путем плавного изменения фокусного расстояния и определения отдельно для каждой точки сцены момента, когда ее локальная резкость достигает максимума. Этот подход называется *дальностью из фокусировки*, поскольку в нем используется информация только о моментах наступления максимальной резкости.

Данный подход не требует детального знания характеристик объектива (вида ядра свертки в зависимости от степени расфокусировки), однако он обладает определенным недостатком, заключающимся в большом времени работы.

Можно заметить, что невозможность использования самих значений резкости в точках расфокусированного изображения связана с тем, что неизвестен пространственный спектр резкого изображения. Возможно, что резкое изображение не содержит тонких деталей, и энергия в его спектре сосредоточена в области низких гармоник, или, наоборот, изображение может содержать мелкозернистую текстуру, которая вносит основной вклад в область высоких гармоник. Расфокусировка изображения второго типа может дать то же распределение энергии в пространственном спектре, которое наблюдается в резком изображении первого типа.

Если бы имелось эталонное резкое изображение, то можно было бы определять дальность до точек сцены по степени их локального размытия, пользуясь только одним расфокусированным изображением. Для получения резкого изображения требуется большая глубина резкости, а для получения расфокусированного изображения (причем, с разным размытием для точек, находящихся на разном удалении) нужна малая глубина резкости. Таким образом, может оказаться необходимым использовать две разные камеры. Вместо этого можно использовать одну камеру, с помощью которой делать несколько снимков с разным фокусным расстоянием (рис. 3.23) и определять расстояние до точек путем сравнения степеней размытия на полученных изображениях.

При этом не требуется плавного изменения фокусного расстояния, а необходимо лишь получение изображений на его нескольких дискретных значениях. Здесь используются не моменты наступления максимальной резкости, а количественные значения степени резкости, поэтому метод называется *дальностью из дефокусировки*. В этом методе нужно получение меньшего числа изображений, чем в методе определения дальности из фокусировки, но при этом требуется более детальная информация о камере (о функции размытия точки при разной степени расфокусировки). Эта информация, как правило, получается в результате достаточно трудоемкого процесса калибровки.



Рис. 3.23. Изображения, полученные с использованием разного фокусного расстояния: разные области изображения имеют разную резкость

### Вопросы и упражнения

1. Как в методе структурированной подсветки определяется дальность до точек сцен? Каковы основные особенности (преимущества и недостатки) этого метода?
2. С помощью каких устройств реализуется метод определения дальности из фазы? Каким образом производится измерение фазы в случае электромагнитного излучения?
3. В чем заключается идея метода восстановления формы из текстуры?
4. При каких условиях с помощью метода формы из текстуры может быть восстановлен наклон плоской поверхности?
5. Как можно использовать текстурную информацию для анализа формы без ее явного восстановления?
6. Чем отличаются методы восстановления дальности из фокусировки и дефокусировки?
7. С помощью каких критериев может оцениваться резкость изображений? Какие особенности дискретного преобразования Фурье нужно учитывать при использовании спектральной информации для оценки резкости?

## 4. Восстановление физических характеристик объектов и понимание изображений

### 4.1. Восстановление отражательной способности

#### *Отражательные свойства поверхностей*

Из обыденного опыта каждый человек знает, что в зависимости от освещения некоторая сцена может выглядеть по-разному. В то же время, при одинаковом освещении по-разному выглядят разные объекты. Таким образом, изображение зависит как от свойства самих объектов, так и от особенностей освещения. Цвет объекта определяется долей отражаемого им света в разных диапазонах длин волн. В случае полутонных изображений важна общая доля отраженного света (в соответствии с кривой чувствительности приемника излучения), называемого альбедо (к этому понятию мы вернемся чуть позже). Если поверхность отражает весь падающий свет, она будет белой; при полном поглощении – черной.

Однако при одних условиях освещения белая поверхность может оказаться менее яркой, чем серая поверхность при других условиях освещения (причем эта ситуация может возникнуть на одном изображении). Тем не менее, человек безошибочно определяет их цвет (см. рис. 4.1). Даже если сцена освещается не белым, а окрашенным светом, человек может достаточно точно идентифицировать «истинные» цвета поверхностей, что было впервые проверено Э. Лэндом экспериментально (им создавались условия, в которых свет, отраженный от листов бумаги разного цвета, при регистрации фотометрическими приборами оказывался идентичным; в то же время люди могли отличить листы по цвету).

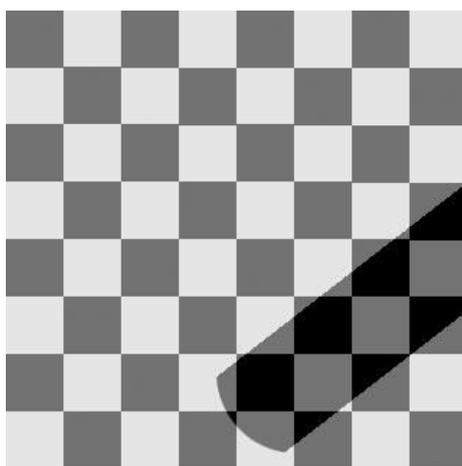


Рис. 4.1. Пример изображения, на котором присутствуют области разной яркости, но им соответствуют поверхности с одинаковой отражательной способностью (белый квадрат в тени обладает той же яркостью, что и серый квадрат снаружи тени, однако воспринимается как более светлый)

Способность воспринимать истинный цвет или отражательную способность поверхности независимо от условий освещения называется *константностью восприятия* или (в случае систем компьютерного зрения) *инвариантностью* по отношению к условиям съемки.

Инвариантность по отношению к условиям освещения обычно достигается выделением на изображении контуров (см. рис. 4.2), о преимуществе использования которых уже говорилось ранее. Однако в некоторых случаях важную информацию несет отражательная способность поверхностей, которую необходимо восстановить в явном виде.

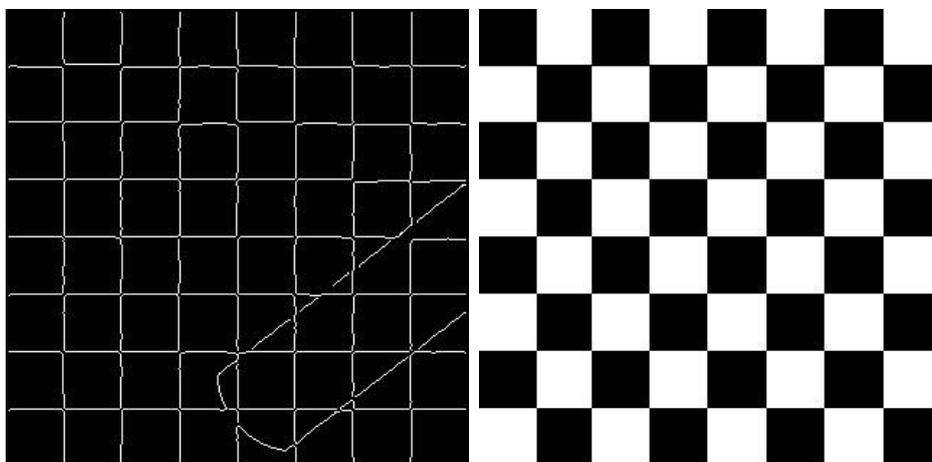


Рис. 4.2. Два возможных пути достижения инвариантности по отношению к условиям освещения: путем выделения контуров и путем восстановления отражательной способности поверхности

Рассмотрим простейший случай, когда видимая яркость точки поверхности определяется произведением освещенности сцены в этой точке и отражательной способности поверхности. Подобные условия можно получить, составив сцену из однородных плоских поверхностей одинаковой ориентации (рис. 4.3). Такую специальную сцену Э. Лэнд назвал *мондрианом* (по имени художника П.К. Мондриана).



Рис. 4.3. Пример мондриана

Даже если освещенность является постоянной для всех точек сцены, проблема может не иметь точного и однозначного решения. Самый простой способ восстановления отражательной способности поверхностей заключается в поиске на изображении области с наибольшей яркостью и

принятии допущения, согласно которому отражательная способность соответствующей поверхности равна 1. Далее альбедо всех остальных поверхностей определяются по отношению их видимых яркостей.

Если сцена зафиксирована и просто меняется мощность источника освещения, то результат работы этого простого метода меняться не будет. Сходным образом можно поступить и в случае цветных изображений по каждому из каналов. Однако если отражательная способность самой яркой поверхности меньше единицы (в некотором цветовом канале), то все цвета будут восстановлены с ошибкой – отражательные способности будут завышены. Напротив, если в поле зрения попадает источник света, отражательные способности поверхностей могут оказаться сильно заниженными.

Для нормировки может применяться не только максимальная яркость точек изображения, но и другие характеристики, например, средняя яркость или какие-то другие статистические моменты распределения значений яркости, которые характеризуют среднюю освещенность сцены. К сожалению, универсальных методов восстановления истинной отражательной способности поверхностей без знания характеристик источников освещения не существует. Видимо, в человеческой зрительной системе для решения этой проблемы одновременно используются разные характеристики изображения (возможно, в том числе и информация о конкретных объектах).

В более реалистичном случае освещенность точек сцены различна. При этом, как правило, освещенность достаточно плавно меняется от точки к точке (за исключением резких границ теней). В то же время изменение отражательной способности поверхностей обычно не носит плавного характера, а на границах объектов изменяется дискретно. Таким образом, информация об изменении отражательной способности преимущественно содержится в области высоких, а об изменении освещенности – в области низких (или средних) частот пространственного спектра. Именно поэтому выделение контуров, основанное на выделении высокочастотной составляющей изображений, приводит к инвариантности по отношению к условиям освещения.

Задача устранения плавных вариаций освещенности с сохранением информации об отражательной способности является более сложной, поскольку она не может быть решена только за счет использования локальных операторов. Интересно отметить, что логарифм яркости будет представлять собой сумму логарифмов освещенности и отражательной способности, одну из которых можно считать гладкой, а другую – ступенчатой функцией. В некотором приближении это может быть сделано, однако на практике ситуация оказывается более сложной, поскольку яркость поверхности плавно меняется с изменением ее ориентации.

### Основные понятия фотометрии

Изображение обычно формируется как результат регистрации интенсивности электромагнитного излучения, приходящего с разных направлений. Это освещение генерируется некоторым источником освещения и при распространении взаимодействует с веществом. И поскольку оно несет в себе результат этого взаимодействия, изображения могут быть использованы для восстановления информации о внешнем мире. Для оптических изображений характерны процессы рассеяния излучения, в то время как (зеркальным) отражением, преломлением и поглощением в системах компьютерного зрения часто пренебрегают, если эти системы не предназначены для решения каких-то особых задач. Очевидно, изображение объекта будет зависеть от его формы, отражательных свойств, а также расположения (и типа) источника освещения и системы, формирующей изображение (рис. 4.4).

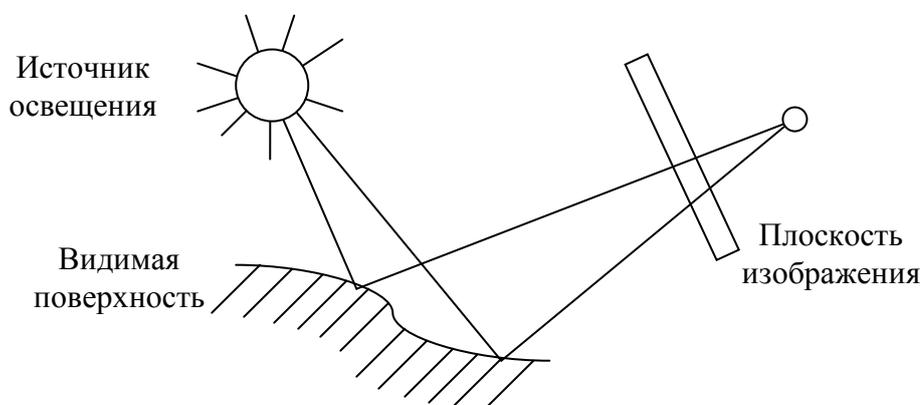


Рис. 4.4. Процесс формирования изображения, в котором к каждой видимой точке поверхности можно однозначно приписать направление на источник освещения и направление на наблюдателя

В связи с этим значения яркостей пикселей может использоваться как для определения формы объекта, так и для определения его отражательных свойств. Для этого, однако, нужно понять, как связаны значения мощности источника света, ориентации поверхности относительно источника и наблюдателя и т.д. Эти вопросы изучаются в *фотометрии*, в которой вводится ряд понятий.

*Освещенностью* поверхности называется мощность излучения, падающего на единицу площади поверхности. Освещенность является интегральной характеристикой, поскольку в ней не учитывается, с какого направления приходит сколько энергии. Если, к примеру, рассмотреть зеркальную поверхность, то даже при неизменной освещенности ее вид будет существенно зависеть от направления на источник освещения. В связи с этим представляет интерес освещенность не объектов сцены, а точек изображения (вернее, освещенность тех физических элементов, которые переводят падающий свет в электрический сигнал или регистрируют свет каким-либо другим образом).

Вместо освещенности объектов сцены удобнее рассматривать понятие яркости, которая определяется как мощность света, испускаемого (излучаемого, рассеиваемого или отражаемого) единицей поверхности в единичном телесном угле в заданном направлении (рис. 4.5). Величина телесного угла определяется как площадь вырезаемого им участка единичной сферы.

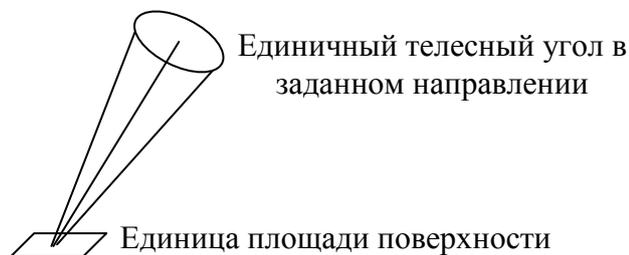


Рис. 4.5. Иллюстрация к понятию яркости поверхности, определяемой мощностью излучения, исходящего с единичной площадки в единичном телесном угле в заданном направлении

Из определения видно, что яркость точек поверхности может зависеть от угла обзора. Между яркостью поверхности и освещенностью площадки изображения существует достаточно простая зависимость:

$$E = \frac{\pi}{4} \left( \frac{d}{f} \right)^2 L \cos^4 \alpha, \quad (4.1)$$

где  $E$  – освещенность площадки изображения,  $L$  – яркость в соответствующем направлении и соответствующем телесном угле; направление  $\alpha$  определяется положением площадки на изображении (см. рис. 4.6), а телесный угол – размером этой площадки и фокусным расстоянием  $f$  и диаметром линзы  $d$  (чем больше размер линзы, тем больше энергии она собирает). Зависимость (4.1) выводится в фотометрии из достаточно простых геометрических соображений, которые здесь повторяться не будут.

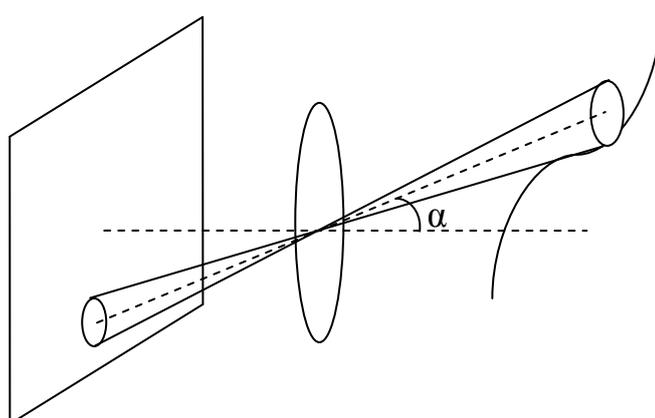


Рис. 4.6. Иллюстрация к связи между освещенностью точки изображения и яркостью поверхности

Принципиальным следствием из уравнения (4.1) является то, что освещенность пикселя в нулевом приближении пропорциональна яркости соответствующей точки сцены (при этом отсутствует зависимость от

дальности до точек сцены). Зависимостью от направления  $\alpha$  можно пренебречь (либо при необходимости компенсировать ее на предварительном этапе обработки изображения). В связи с этим, измеренную величину освещенности в каждом пикселе также называют яркостью (поскольку абсолютная мощность источников освещения обычно неизвестна, то коэффициент пропорциональности между яркостями пикселей и яркостями точек сцены роли не играет).

Посмотрим теперь, от чего зависит яркость поверхности (и, соответственно, яркость пикселей). Как уже отмечалось, яркость поверхности зависит от направления на наблюдателя и на источник света. Существует интегральная характеристика коэффициента отражения, *альbedo*, показывающая, какая доля падающего на поверхность света отражается. Однако нас в данном случае интересует более детальная характеристика – отражательная способность, определяемая как доля отражаемого (или рассеиваемого) света в разных направлениях в зависимости от угла падения света.

Для описания всех возможных направлений, под которым может падать свет на элемент поверхности, требуется два угла – угол между направлением падающего света и нормалью к поверхности (полярный угол) и азимутальный угол (так же как для описания положения звезды на небе требуется ее высота над горизонтом и азимут направления). Также два угла требуются для описания направления рассеяния. Поскольку отражательная способность зависит от двух направлений, она обычно описывается с помощью *двулучевой функции отражательной способности (ДФОС)*  $A(\varphi_i, \lambda_i, \varphi_e, \lambda_e)$ , где  $(\varphi_i, \lambda_i)$  – полярный угол и азимут направления освещения, а  $(\varphi_e, \lambda_e)$  – полярный угол и азимут направления на наблюдателя.

Для большинства поверхностей (за исключением поверхностей материалов с анизотропными свойствами) ДФОС не меняется при вращении поверхности вокруг нормали. При этом, однако, зависимость от разности азимутов  $\lambda_i - \lambda_e$  остается, что хорошо видно на примере зеркальных поверхностей (см. также рис. 4.7).

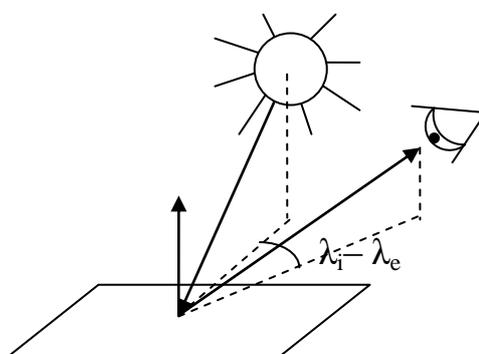


Рис. 4.7. Иллюстрация к наличию зависимости от разницы азимутальных углов направлений на источник освещения и на наблюдателя

Если имеется протяженный источник освещения, то, очевидно, яркость точки поверхности будет определяться путем интегрирования

$$L(\varphi_e, \lambda_e) = \int_{-\pi}^{\pi} \int_0^{\pi/2} A(\varphi_i, \lambda_i, \varphi_e, \lambda_e) E(\varphi_i, \lambda_i) \sin \varphi_i \cos \lambda_i \, d\varphi_i \, d\lambda_i,$$

где  $E(\varphi_i, \lambda_i)$  – мощность излучения, приходящего на площадку поверхности от источника света по направлению  $(\varphi_i, \lambda_i)$  в единице телесного угла.

Для точечного источника, располагающегося в направлении  $(\varphi_i, \lambda_i)$  относительно локальной площадки поверхности, мощность излучения может быть выражена в форме  $E(\varphi_i, \lambda_i) = E\delta(\varphi - \varphi_i)\delta(\lambda - \lambda_i)$ , где  $\delta$  – дельта-функция.

Другой крайний случай – это полностью рассеянное излучение, для которого  $E(\varphi_i, \lambda_i) = \text{const}$ . Очевидно, зависимость яркости поверхности от ее ориентации относительно источника света будет отсутствовать (но сохранится зависимость от ориентации относительно наблюдателя).

Для поверхности произвольного типа ДФОС имеет почти произвольный вид, что затрудняет ее использование на практике. Она может быть определена экспериментально путем построения изображения некоторой поверхности при поочередном освещении с разных сторон, однако в компьютерном зрении это используется крайне редко. Вместо этого обычно рассматриваются частные типы поверхностей.

Наиболее широко распространенным приближением ко многим реальным поверхностям является *идеальная ламбертова поверхность*, которая рассеивает свет одинаково во всех направлениях. Этот идеализированный случай достаточно хорошо описывает матовые поверхности. ДФОС такой поверхности будет константной. В простейшем случае полагается, что ламбертова поверхность рассеивает весь падающий свет, ничего не поглощая, тогда несложно убедиться, что ее ДФОС будет иметь вид

$$A(\varphi_i, \lambda_i, \varphi_e, \lambda_e) = 1/\pi.$$

Если же часть света поглощается, то необходимо учитывать полное альбедо (долю рассеянного света  $A_0$ ), и тогда получим

$$A(\varphi_i, \lambda_i, \varphi_e, \lambda_e) = A_0/\pi. \quad (4.2)$$

Яркость ламбертовой поверхности, выводимая в фотометрии, в случае точечного источника, располагающегося по направлению  $(\varphi_i, \lambda_i)$  и имеющего мощность  $E$ , принимает вид

$$L = \frac{1}{\pi} A_0 E \cos \varphi_i \quad (\text{при } \varphi_i \geq 0), \quad (4.3)$$

что представляет собой *закон Ламберта* для матовых поверхностей, или закон косинуса.

Падение яркости пропорционально косинусу полярного угла связано с уменьшением освещенности поверхности при ее наклоне относительно направления на источник света, так как при этом видимая (со стороны

источника освещения) площадь поверхности уменьшается пропорционально  $\cos\varphi_i$ .

Уменьшение яркости точек поверхности с увеличением наклона носит название *затенения*, хотя поверхность и не находится в области тени, что может внести определенную путаницу при использовании термина.

Таким образом, яркость матовой поверхности зависит от направления на источник света и не зависит от направления на наблюдателя. К примеру, если взять лист белой бумаги и смотреть на него с разных сторон, то его яркость будет неизменной. Если же вращать этот лист, то его яркость будет меняться (так как при этом будет меняться не только направление на наблюдателя, но и на источник освещения). Очевидно, при рассеянном освещении зависимость яркости поверхности от направления на источник освещения пропадет.

Существует другой частный случай – случай зеркальной поверхности, для которой ДФОС можно получить в виде

$$A(\varphi_i, \lambda_i, \varphi_e, \lambda_e) = A_0 \frac{\delta(\varphi_e - \varphi_i) \delta(\lambda_e - \lambda_i - \pi)}{\sin \varphi_i \cos \varphi_i}.$$

Если определить яркость точек зеркальной поверхности, то можно получить, что они будут просто соответствовать зеркальному отражению источника:

$$L(\varphi_e, \lambda_e) = E(\varphi_e, \lambda_e - \pi).$$

Промежуточное место между зеркальными и матовыми поверхностями занимают глянцевые поверхности, которые рассеивают значительную часть света в направлении отражения, что принимает форму бликов, когда  $\varphi_e = \varphi_i$  и  $\lambda_e = \lambda_i + \pi$ . Такие поверхности достаточно часто встречаются на практике (рис. 4.8), и для них в компьютерном зрении возникает отдельная задача фильтрации бликов.



Рис. 4.8. Примеры появления бликов

Для глянцевых поверхностей, как и для матовых, характерен эффект затенения. Поскольку затенение (понижение яркости) вызвано изменением ориентации поверхности, то этот эффект может быть использован для восстановления трехмерной формы объектов.

### Фотометрическое стерео

Наиболее непосредственным, хотя и применимым на практике в достаточно малом числе случаев, методом использования информации о затенении является метод *фотометрического стерео*. Этот метод заключается в получении двух изображений статичной сцены при неизменном положении камеры, но разных (причем известных) положениях источника освещения.

Рассмотрим случай ламбертовой поверхности и будем считать, что источник, достаточно удален от сцены (такие источники освещают сцену плоскопараллельным потоком излучения, и направление, по которому приходит излучение, не зависит от координат точки на сцене).

Рассмотрим единичную площадку. Для нее будут иметься два значения яркости

$$L_1 = \frac{1}{\pi} A_0 E \cos \varphi_{i,1} \quad \text{и} \quad L_2 = \frac{1}{\pi} A_0 E \cos \varphi_{i,2}.$$

Здесь полярные углы  $\varphi_{i,1}$  и  $\varphi_{i,2}$  являются углами между направлениями на положения источника освещения и нормалью к поверхности.

Рассмотрим некоторую абсолютную систему координат, в которой нормаль к наблюдаемой поверхности в данной точке описывается неизвестным единичным вектором  $\mathbf{n} = (n_x, n_y, n_z)$ , а два направления на источники освещения – единичными векторами  $\mathbf{v}_1 = (v_{x,1}, v_{y,1}, v_{z,1})$  и  $\mathbf{v}_2 = (v_{x,2}, v_{y,2}, v_{z,2})$ , причем эти два вектора известны. Несложно убедиться, что  $\cos \varphi_{i,1} = \mathbf{v}_1 \mathbf{n} = v_{x,1} n_x + v_{y,1} n_y + v_{z,1} n_z$ , то есть скалярному произведению соответствующих векторов (см. рис. 4.9).

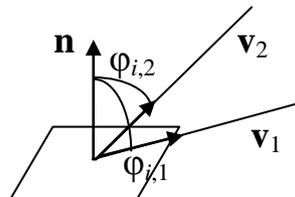


Рис. 4.9. Иллюстрация к определению полярных углов через произведения векторов нормали и направления на источники освещения в задаче фотометрического стерео

Также получаем  $\cos \varphi_{i,2} = \mathbf{v}_2 \mathbf{n} = v_{x,2} n_x + v_{y,2} n_y + v_{z,2} n_z$ . Таким образом, имеем систему из двух уравнений:

$$\begin{cases} v_{x,1} n_x + v_{y,1} n_y + v_{z,1} n_z = \frac{L_1}{A_0 E \pi} \\ v_{x,2} n_x + v_{y,2} n_y + v_{z,2} n_z = \frac{L_2}{A_0 E \pi} \end{cases}$$

Если величины  $E$  и  $A_0$  считать известными, то данная система состоит из двух уравнений и содержит две независимые переменные, поскольку вектор  $\mathbf{n}$  является единичным. Если же эти величины являются

неизвестными, то можно рассмотреть вектор  $\pi A_0 E \mathbf{n}$ , который уже не будет являться единичным. Для его восстановления потребуется три уравнения, полученных при разных направлениях на источник освещения.

Таким образом, в методе фотометрического стерео достаточно использовать локальную информацию с двух или трех изображений для восстановления вектора нормали к поверхности в каждой ее точке. Естественно, задача имеет решение только в невырожденных случаях, например, когда три направления на различные положения источника освещения не являются компланарными (то есть источник света не должен перемещаться в одной плоскости, что имеет место, в частности, в случае Солнца).

### *Форма из затенения*

В большинстве задач положение источников освещения априори неизвестно. Тем не менее, и в этих условиях зрительная система человека эффективно использует информацию о затенении для восстановления трехмерной формы объектов.

В этом случае локальное решение проблемы невозможно, и необходимо использовать информацию о затенении одновременно в разных точках. Рассмотрим случай ламбертовой поверхности с постоянной отражательной способностью:

$$L(x, y) = \frac{1}{\pi} A_0 E \cos \varphi_i(x, y).$$

Здесь полярный угол зависит от ориентации поверхности в каждой точке, что можно представить как зависимость от координат точки.

Будем полагать, что направление прихода излучения  $\mathbf{v}$  одинаково для всех точек сцены, тогда полярный угол будет определяться только локальной ориентацией поверхности:

$$\cos \varphi_i(x, y) = \mathbf{v} \mathbf{n}(x, y) = v_x n_x(x, y) + v_y n_y(x, y) + v_z n_z(x, y),$$

откуда имеем  $L(x, y) = \frac{1}{\pi} A_0 E (v_x n_x(x, y) + v_y n_y(x, y) + v_z n_z(x, y))$ , причем

$$n_z(x, y) = \sqrt{1 - n_x^2(x, y) - n_y^2(x, y)}.$$

Это может быть неверно для близкорасположенного источника освещения. К примеру, если источник располагается над сферой, то полярный угол будет быстрее увеличиваться к краям сферы по сравнению со случаем удаленного источника. Следовательно, яркость сферы у краев будет ниже.

Решение данной задачи сложнее, чем задачи фотометрического стерео. Опишем лишь ряд общих идей. Часто на поверхности мы можем найти точку наибольшей яркости, для которой векторы  $\mathbf{v}$  и  $\mathbf{n}$  являются коллинеарными. Поскольку мы можем произвольным образом выбирать систему координат, в которой описывается форма объекта, вектор нормали в точке максимальной яркости можно задать, например, как  $\mathbf{n}(x^*, y^*) = (0, 0, 1)$ . Можно полагать, что

в этой точке  $\cos\varphi_i(x^*, y^*) = 1$ . Для любой точки поверхности можно найти косинус полярного угла как  $L(x, y)/L(x^*, y^*)$ . Для уменьшения влияния шумов следует предположить гладкость нормали к поверхности как функции от координат ее точек. Тогда функцию  $p(x, y) = \cos\varphi_i(x, y)$  можно найти, минимизируя функционал

$$\iint \left( \left[ p(x, y) - L(x, y)/L(x^*, y^*) \right]^2 + |\nabla p(x, y)|^2 \right) dx dy.$$

Построим теперь на изображении линии, состоящие из точек с постоянным полярным углом. В каждой точке таких линий можно построить перпендикуляр к ним, который будет соответствовать азимутальному углу. В качестве решения может выступать любой из двух перпендикуляров. Поскольку изменение азимутального угла является непрерывным, выбор одного из двух направлений должен осуществляться одновременно для всех точек поверхности. Эти два решения соответствуют вогнутой и выпуклой поверхностям, которые только по данным о затенении не различаются.

Представленные здесь соображения являются лишь общими идеями восстановления формы из затенения. Часто используются более сложные методы, основанные на оценивании локальной кривизны поверхности и ее интеграции вдоль характеристических линий для получения полного решения. Однако на практике эти методы также имеют ограниченное применение в силу того, что поверхности с постоянной отражательной способностью встречаются крайне редко. Кроме того, понижение яркости точек поверхности может быть обусловлено не только затенением, но и тенями, которые представляют собой отдельную трудность при анализе изображений, но также несут информацию о форме объектов.

### *Высота из тени*

Наиболее надежным использование тени оказывается для аэрокосмических изображений, поскольку они отбрасываются на сравнительно плоскую поверхность и обусловлены одним бесконечно удаленным источником освещения. В связи с этим тени на аэрокосмических изображениях обладают двумя основными особенностями, по которым они могут быть идентифицированы, – это низкая яркость и одинаковая ориентация (рис. 4.10).

На подобных изображениях границы теней, как правило, достаточно точно повторяют границы отбрасывающих их объектов. Путем выполнения корреляционного или чамферного сопоставления границ теней и объектов позволяет найти смещение между ними в плоскости изображения, и это смещение пропорционально высоте объектов. Часто при съемке аэрокосмических изображений известна ориентация камеры и положение Солнца, благодаря чему по длине тени могут быть восстановлены абсолютные значения высот объектов.



Рис. 4.10. Пример аэрокосмического изображения с тенями, которые могут использоваться для оценивания высоты зданий

Использование информации о тенях на изображениях сцен внутри помещений также возможно, хотя и заметно более затруднительно. На настоящий момент тени для изображений этого типа чаще рассматриваются в качестве мешающего фактора, хотя человек их и здесь может идентифицировать и использовать для анализа содержания сцены.

### Вопросы и упражнения

1. Что такое константность восприятия? Какое отношение она имеет к проблемам компьютерного зрения?
2. Как связана освещенность точек сцены с яркостью соответствующих пикселей изображения?
3. Что такое двулучевая функция отражательной способности?
4. В чем состоит закон Ламберта? Как он используется при построении систем компьютерного зрения?
5. Что подразумевается под задачей фотометрического стерео?
6. При каких условиях может быть восстановлена информация о локальной ориентации поверхности по ее яркости?
7. В чем состоит отличие матовых, глянцевых и зеркальных поверхностей?
8. Какие ограничения имеют методы восстановления трехмерной формы из затенения?

## 4.2. Анализ движения

Помимо трехмерной формы и пространственного расположения объектов, а также отражательных свойств их поверхностей в результате анализа изображений может восстанавливаться информация о движении объектов. Для этого, естественно, необходимо иметь последовательность кадров, полученных в разные моменты времени, или видеоряд. Кроме того, использование видеоряда позволяет надежнее решать и другие задачи анализа изображений и компьютерного зрения (например, сегментации изображений на объекты, восстановления трехмерных координат точек сцены и т.д.).

Сформулируем проблему анализа движений по видеоряду. Для этого заметим, что каждая точка сцены обладает некоторой скоростью в системе координат камеры. Проекция этой точки также будет обладать некоторой скоростью в плоскости изображения. Несложно найти связь между этими скоростями, исходя из уравнения центральной проекции:

$$x'/f' = x/z \text{ и } y'/f' = y/z.$$

Через момент времени  $dt$  связь координат примет вид

$$\begin{aligned} (x' + v'_x dt)/f' &= (x + v_x dt)/(z + v_z dt), \\ (y' + v'_y dt)/f' &= (y + v_y dt)/(z + v_z dt), \end{aligned}$$

где  $v'_x$  и  $v'_y$  – скорости проекции точки в плоскости изображения,  $v_x, v_y$  и  $v_z$  – скорости точки в физическом пространстве.

Преобразуем первое уравнение, пренебрегая слагаемым с множителем  $dt^2$ :

$$x'z + zv'_x dt + x'v_z dt = f'x + f'v_x dt$$

Учитывая, что  $x'z = f'x$ , имеем

$$zv'_x + x'v_z = f'v_x.$$

Таким образом, можно получить

$$\begin{cases} v'_x = (f'v_x - x'v_z)/z, \\ v'_y = (f'v_y - y'v_z)/z. \end{cases} \quad (4.4)$$

Итак, каждой точке изображения можно приписать некоторый вектор скорости, являющийся проекцией вектора скорости соответствующей физической точки. Совокупность векторов скоростей всех точек изображения называется *полем движения*, которое, как правило, априорно неизвестно. Проблема восстановления поля движения является проблемой анализа изображений. К области компьютерного зрения можно отнести проблему восстановления вектора физических скоростей  $(v_x, v_y, v_z)$ .

Как несложно увидеть из уравнений (4.4), в произвольном случае по полю движения однозначно восстановить физические скорости, к сожалению, невозможно. Кроме того, даже поле скоростей далеко не всегда можно корректно определить по видеоряду.

К примеру, представим себе однородную матовую сферу. Вращение этой сферы не приведет к изменению изображения. Напротив, изображение статичной сферы будет меняться при перемещении источника освещения. При этом будет видно кажущееся вращение сферы вслед за изменением распределения яркости по ее поверхности. В связи с тем, что видимое перемещение точек изображения может не соответствовать реальному полю движения, для обозначения видимого (кажущегося) движения яркостной картинки вводят дополнительный термин – *оптический поток*. В частности, оптический поток для изображения вращающейся однородной матовой сферы является нулевым в отличие от поля движения, а для изображения статичной сферы при движущемся источнике освещения ситуация оказывается обратной.

Оптический поток совпадает с полем движения лишь в идеальных условиях, однако при некоторых ограничениях он может служить неплохой аппроксимацией поля движения, поэтому часто говорят о задаче оценивания оптического потока как базовой задаче анализа движения.

Как несложно заметить, проблему анализа движения можно трактовать как специфическую проблему сопоставления изображений. Действительно, для оценки поля движения или оптического потока необходимо отождествить сопряженные точки на соседних кадрах видеоряда. Как и любая задача сопоставления изображений, данная задача может решаться с привлечением представлений различного уровня абстракции. Поскольку оптический поток, по определению, затрагивает все точки на изображении и соответствует видимому движению яркостного изображения, естественным является оценивание оптического потока путем отождествления точек одинаковой яркости (то есть с использованием низкоуровневых представлений).

#### *Оценивание оптического потока*

Пусть изображение задано как изменяющееся во времени поле яркости  $f(x, y, t)$ . В пространственно-временной окрестности точки  $(x, y, t)$  это поле можно разложить в ряд Тейлора:

$$f(x + \Delta x, y + \Delta y, t + \Delta t) = f(x, y, t) + \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y + \frac{\partial f}{\partial t} \Delta t + \varepsilon,$$

где  $\varepsilon$  – сумма слагаемых, содержащих частные производные вторых и более высоких степеней. При малых изменениях  $\Delta x, \Delta y, \Delta t$  можно полагать  $\varepsilon \approx 0$ . Это допустимо при условии, что точки изображения на соседних кадрах смещаются не более чем на 1 пиксель. Если  $\Delta x$  и  $\Delta y$  – это смещения точки  $(x, y)$  за время  $\Delta t$ , то в некотором приближении верно равенство  $f(x + \Delta x, y + \Delta y, t + \Delta t) = f(x, y, t)$ . Откуда получаем условие

$$\frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y + \frac{\partial f}{\partial t} \Delta t = 0,$$

которое можно преобразовать к виду

$$\frac{\partial f(x, y, t)}{\partial x} v_x(x, y, t) + \frac{\partial f(x, y, t)}{\partial y} v_y(x, y, t) + \frac{\partial f(x, y, t)}{\partial t} = 0, \quad (4.5)$$

где  $v_x(x, y) = \Delta x / \Delta t$  и  $v_y(x, y) = \Delta y / \Delta t$  – оценки скоростей перемещения точек на изображении. Уравнение (4.5) называется *уравнением оптического потока*.

В этом уравнении мы можем для каждого момента времени и каждой точки изображения оценить все частные производные  $\left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial t} \right)$ . Для оценки частной производной по времени необходимо иметь, по крайней мере, два кадра видеоряда.

Таким образом, видно, что для каждой точки изображения одно уравнение оптического потока связывает две неизвестные  $v_x$  и  $v_y$ , которые однозначно восстановить нельзя. Восстанавливается лишь проекция вектора скорости на направление пространственного градиента яркости  $\left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$ .

Действительно, несложно убедиться, что если пара значений  $v_x^{(0)}$  и  $v_y^{(0)}$  являются решением уравнения оптического потока (4.5), то пара  $v_x^{(0)} - \alpha \frac{\partial f}{\partial y}$  и

$v_y^{(0)} + \alpha \frac{\partial f}{\partial x}$  также будет решением при любом значении  $\alpha$ .

Это можно наглядно проиллюстрировать, построив на изображении линии постоянной яркости. Величина перемещения может быть определена только в направлении, перпендикулярном направлению линии, но не вдоль нее. К примеру, на рис. 4.11а представлены кадры видеоряда, по которому невозможно установить компонент вращения фигуры вокруг своей оси. На рис. 4.11б приведены два положения движущейся линии, в каждой точке которой градиент яркости направлен горизонтально, в связи с чем может быть найдена только горизонтальная компонента вектора перемещения.

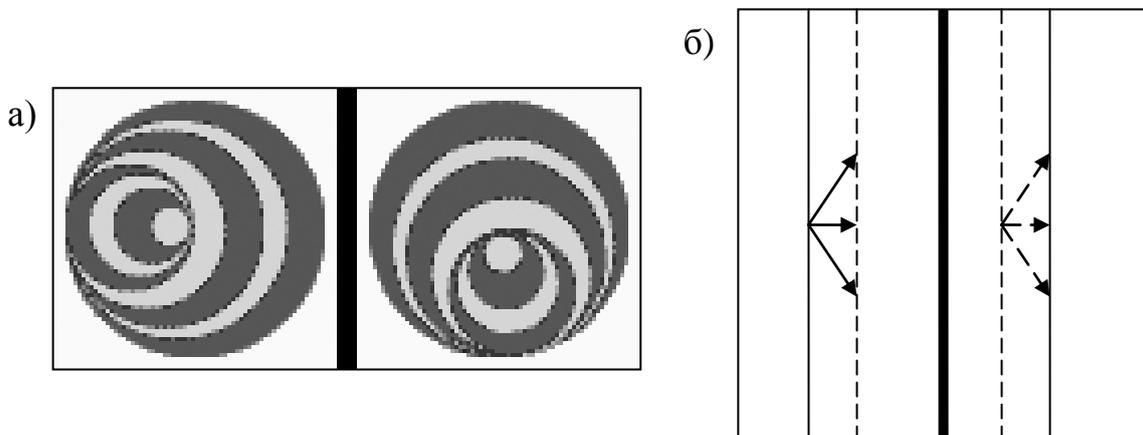


Рис. 4.11. Примеры невозможности определения компоненты оптического потока в направлении, перпендикулярном направлению пространственного градиента яркости

Кроме того, если точка изображения располагается внутри области постоянной яркости, то все три частные производные (по  $x$ ,  $y$  и  $t$ ) будут равны нулю. Следовательно, для таких точек вообще никакая информация о направлении движения не будет доступна.

Таким образом, определение оптического потока с использованием уравнения (4.5) сопряжено с существенной неоднозначностью. Наличие шумов на изображении приводит к заметному искажению оцененных значений скоростей.

Уменьшение неоднозначности и влияния шумов может быть достигнуто за счет построения оптического потока для всех точек одновременно с введением ряда дополнительных предположений. В рамках функциональных моделей изображений для этого вводится некоторый минимизируемый функционал. В качестве слагаемого этот функционал должен содержать интегральную невязку уравнения оптического потока

$$\varepsilon_o = \sum_x \sum_y \left[ \frac{\partial f(x, y, t)}{\partial x} v_x(x, y, t) + \frac{\partial f(x, y, t)}{\partial y} v_y(x, y, t) + \frac{\partial f(x, y, t)}{\partial t} \right]^2.$$

Эта невязка, однако, может минимизироваться для каждой точки независимо.

Однако для реальных сцен скорости в соседних точках изображения не являются независимыми. Согласно одному из предположений Д. Марра, рассмотренных ранее, оптический поток почти всюду гладок. Это условие говорит о том, что значения пространственных частных производных вектора скоростей  $(v_x(x, y, t), v_y(x, y, t))$  должны быть малы. Это позволяет ввести еще одно слагаемое в минимизируемый функционал:

$$\varepsilon_g = \sum_x \sum_y \left[ \left( \frac{\partial v_x(x, y, t)}{\partial x} \right)^2 + \left( \frac{\partial v_x(x, y, t)}{\partial y} \right)^2 + \left( \frac{\partial v_y(x, y, t)}{\partial x} \right)^2 + \left( \frac{\partial v_y(x, y, t)}{\partial y} \right)^2 \right].$$

Иногда вводится еще одно дополнительное предположение, согласно которому сами скорости должны быть не слишком большими, что соответствует введению нового слагаемого:

$$\varepsilon_v = \sum_x \sum_y [v_x^2(x, y, t) + v_y^2(x, y, t)].$$

Минимизируемый функционал при этом представляется в виде суммы:

$$L = a_o \varepsilon_o + a_g \varepsilon_g + a_v \varepsilon_v, \quad (4.6)$$

где  $a_{o,g,v}$  – некоторые коэффициенты, подбираемые в зависимости от особенностей изображений (зашумленности, движущихся объектов и т.д.).

Алгоритмы минимизации этого функционала являются вычислительно затратными и дают не слишком надежное решение задачи восстановления оптического потока.

В рамках стохастических моделей изображений восстановление оптического потока производится сходным образом, хотя и с использованием другого математического аппарата. Здесь рассматривается плотность распределения вероятностей  $p(f_1, f_2 | \chi)$ , где два изображения  $f_1$  и  $f_2$

получены последовательно, а  $\chi$  – параметры некоторой модели движения (в данном случае совокупность значений  $(v_x(x, y, t), v_y(x, y, t))$  для всех точек изображения).

В соответствии с правилом Байеса, апостериорные вероятности значений параметров движения  $\chi$  будут определяться по формуле

$$p(\chi | f_1, f_2) \propto p(\chi) p(f_1, f_2 | \chi).$$

В простейшем случае правдоподобие изображений может быть определено в соответствии с предположением о нормальном законе распределения ошибок:

Часто ограничиваются следующей оценкой правдоподобия для одной точки изображения:

$$p(f_1, f_2 | v_x, v_y) = \alpha \exp \left[ - \frac{\left( \frac{\partial f}{\partial x} v_x + \frac{\partial f}{\partial y} v_y + \frac{\partial f}{\partial t} \right)^2}{2\sigma^2} \right], \quad (4.7)$$

где  $\alpha$  – нормирующий множитель нормального распределения,  $\sigma$  – ожидаемая дисперсия шума.

Произведение значений  $p(f_1, f_2 | v_x, v_y)$  для всех точек изображения даст под экспонентой сумму невязок  $a_o$ , что показывает тесную связь между вероятностными и функциональными моделями.

Априорное распределение вероятностей  $p(\chi)$  может задаваться константным, что будет соответствовать (в совокупности с уравнением (4.7)) просто уравнению оптического потока.

Предположения о малости модулей скоростей и гладкости их изменения от точки к точке могут быть выражены путем задания релевантного распределения  $p(\chi)$ :

$$p(\chi) = \alpha \exp \frac{-[v_x^2(x, y, t) + v_y^2(x, y, t)]}{2\sigma_v^2} \times \\ \times \exp \frac{-\left[ \left( \frac{\partial v_x(x, y, t)}{\partial x} \right)^2 + \left( \frac{\partial v_x(x, y, t)}{\partial y} \right)^2 + \left( \frac{\partial v_y(x, y, t)}{\partial x} \right)^2 + \left( \frac{\partial v_y(x, y, t)}{\partial y} \right)^2 \right]}{2\sigma_g^2}.$$

Если такое априорное распределение не задавать, то результирующие скорости будут большими (и, вообще говоря, будет присутствовать значительная неопределенность в выборе скорости для каждой конкретной точки), а в поле скоростей будет значительное число разрывов.

Здесь видно, что коэффициенты перед слагаемыми в уравнении минимизируемого функционала (4.6) определяются математическими ожиданиями соответствующих величин. В целом, оба низкоуровневых

подхода (на основе функциональных и стохастических моделей) обладают одними и теми же недостатками.

Для увеличения допустимых смещений точек может применяться сглаживание при вычислении частных производных от яркостного изображения. Это, однако, не решает проблемы невозможности получения оценок скоростей точек, находящихся внутри однородных областей, а также оценок компонентов скоростей по направлению, перпендикулярному градиенту яркости. Таким образом, оказывается необходимым доопределять поле скоростей.

Доопределение оптического потока может быть сделано путем интерполяции измеренных значений. Измерение значений скорости окажется наиболее надежным в точках с максимальным модулем градиента яркости, то есть в контурных точках. В точках контура будет определяться только компонента скорости в направлении, перпендикулярном контуру. Однако контур, как правило, имеет изломы. Углы, располагающиеся на контуре, имеют в своей окрестности точки с разным направлением градиента яркости, а значит, для них несложно восстановить полный вектор скорости, пользуясь предположением непрерывности поля скоростей (рис. 4.12).

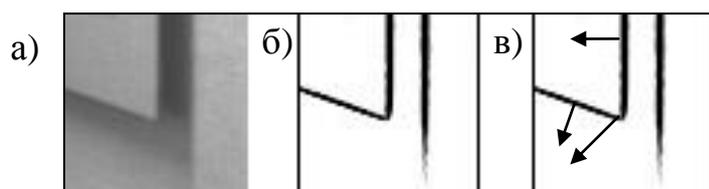


Рис. 4.12. Определение полного вектора скорости в угле путем интерполяции компонентов скорости с соседних контурных точек, обладающих разным направлением градиента: (а) – фрагмент исходного изображения; (б) – краевые точки; (в) – компоненты скорости на прямых участках контура и их объединение в угле

Значения полных векторов скоростей в углах могут быть дальше распространены на все точки контуров и далее – на все точки изображения. Это распространение, однако, является не вполне тривиальным. Следует иметь в виду, что оптический поток не является всюду гладким. В нем могут присутствовать разрывы на границах объектов. Совместно с задачей доопределения оптического потока возникает задача его сегментации, то есть выделения областей, в каждой из которых оптический поток является гладким. Эта задача в определенной степени сходна с задачей текстурной сегментации изображений, однако отличается от нее тем, что не во всех точках значения скоростей определены.

В то же время, на практике далеко не всегда требуется построение оптического потока для всех точек изображения. Весьма часто бывает достаточным его определение только для ряда ключевых точек.

#### *Отождествление сопряженных точек*

Рассмотренный выше способ оценивания оптического потока непосредственно применим только при малых перемещениях точек на

соседних кадрах, поскольку даже при извлечении контуров и обнаружении на них углов (что соответствует уже не низкоуровневому, а контурным и структурным представлениям) оценка скорости получается из уравнения оптического потока, опирающегося на яркостные характеристики.

При значительных перемещениях в рамках низкоуровневого подхода можно было бы строить изолинии (линии постоянной яркости) и определять, куда произошло их перемещение. Однако несложно убедиться, что при вращении объекта яркость его точек будет меняться (за исключением случая рассеянного освещения, который на практике бывает не так часто). Таким образом, на разных кадрах видеоряда изолинии с одинаковой яркостью не будут состоять из сопряженных точек.

Поскольку яркость точек не является хорошей характеристикой для их отождествления, необходимо привлекать представления более высоких уровней. Вместо изолиний можно использовать контуры.

При отождествлении точек контуров на изображениях видеоряда используются те же предположения, что и в низкоуровневых методах оценивания оптического потока. Во-первых, перемещения полагаются не слишком большими. Во-вторых, считается, что оптический поток имеет мало разрывов. Это говорит о том, что поиск сопряженной точки должен осуществляться в окрестности текущей точки, причем для соседних контурных точек сдвиги должны быть близкими. Иными словами, можно выполнять чамферное сопоставление каждого фрагмента контура первого изображения с картой расстояний, построенной на основе контурных точек второго изображения.

Очевидно, если фрагмент соответствует прямолинейному участку контура, то мера Хаусдорфа не будет меняться при перемещениях вдоль контура. Таким образом, и здесь полный вектор скорости может быть определен только для углов. Отличие от рассмотренного ранее случая заключается в способе оценивания скорости. Здесь вместо уравнения оптического потока используется непосредственная оценка перемещения фрагмента контура.

При простейшей реализации данного подхода к оцениванию оптического потока выделяются углы на последовательных изображениях видеоряда, и эти углы ставятся друг другу в соответствие на основе близости их положений. При малых перемещениях такой подход может давать приемлемые результаты, однако при увеличении перемещений отождествление угловых точек только на основе их близости становится невозможным. Дополнительная трудность связана с тем, что некоторые угловые точки могут пропадать или, напротив, появляться, так что для них нет сопряженных точек среди углов, выделенных на другом изображении.

Для увеличения надежности сопоставления могут применяться различные приемы. В рамках признакового подхода каждая ключевая точка описывается вектором инвариантных признаков, на основе которых и производится ее отождествление. Здесь могут применяться те же признаки, что и при сопоставлении изображений трехмерных сцен.

### *Движение твердого тела*

Достаточно распространенным случаем является наблюдение движения твердого тела. Это условие накладывает существенные ограничения на возможные значения перемещений разных точек. Параметры движения твердого тела определяются шестью величинами (три сдвига и три угла поворота), как и параметры движения камеры. В действительности, движение одного твердого тела в поле зрения камеры эквивалентно движению камеры внутри статичной сцены. Несложно заметить, что задача оценивания параметров движения твердого тела (или камеры внутри статичной сцены) практически полностью совпадает с задачей сопоставления изображений трехмерных сцен, снятых с разных ракурсов.

Есть, однако, несколько отличий. Во-первых, в случае задач анализа движения ракурс меняется незначительно, что может быть использовано для построения более эффективных методов сопоставления. Во-вторых, может быть использовано несколько кадров, полученных через равные промежутки времени. Можно полагать, что скорости точек постоянны или меняются не слишком быстро, так что они почти не изменяются для нескольких соседних кадров видеоряда. Использование нескольких кадров позволяет уменьшить число точек, необходимых для восстановления параметров движения. В-третьих, задача оценивания параметров движения твердого тела может быть поставлена не только для центральной, но и для слабоперспективной или ортогональной проекции. Казалось бы, поскольку координаты проекции точки не зависят от ее дальности в случае слабоперспективной проекции, определение трехмерной формы по движению должно оказаться невозможным. Однако это не так, поскольку при вращении объекта формулы для изменения координат  $(x, y)$  его точек будут содержать зависимость от координаты  $z$ .

Рассмотрим кратко частный случай ортогональной проекции, в котором производится измерение положения одной точки на нескольких последовательных кадрах. Будем считать, что изменение трехмерных координат точки описывается одними и теми же параметрами движения вида:

$$\mathbf{r}^{(i+1)} = \mathbf{R}\mathbf{r}^{(i)} + \mathbf{r}_0.$$

Запишем уравнения подробнее для нескольких последующих кадров

$$1) \quad x^{(1)} = r_{11}x^{(0)} + r_{12}y^{(0)} + r_{13}z^{(0)} + \Delta x$$

$$2) \quad y^{(1)} = r_{21}x^{(0)} + r_{22}y^{(0)} + r_{23}z^{(0)} + \Delta y$$

$$3) \quad z^{(1)} = r_{31}x^{(0)} + r_{32}y^{(0)} + r_{33}z^{(0)} + \Delta z$$

$$4) \quad x^{(2)} = r_{11}x^{(1)} + r_{12}y^{(1)} + r_{13}z^{(1)} + \Delta x = \\ = r_{11}x^{(1)} + r_{12}y^{(1)} + r_{13}(r_{31}x^{(0)} + r_{32}y^{(0)} + r_{33}z^{(0)} + \Delta z) + \Delta x$$

$$\begin{aligned}
5) \quad y^{(2)} &= r_{21}x^{(1)} + r_{22}y^{(1)} + r_{23}z^{(1)} + \Delta y = \\
&= r_{21}x^{(1)} + r_{22}y^{(1)} + r_{23}(r_{31}x^{(0)} + r_{32}y^{(0)} + r_{33}z^{(0)} + \Delta z) + \Delta y \\
6) \quad z^{(2)} &= r_{31}x^{(1)} + r_{32}y^{(1)} + r_{33}z^{(1)} + \Delta z = \\
&= r_{31}x^{(1)} + r_{32}y^{(1)} + r_{33}(r_{31}x^{(0)} + r_{32}y^{(0)} + r_{33}z^{(0)} + \Delta z) + \Delta z
\end{aligned}$$

...

Здесь координаты  $x^{(i)}, y^{(i)}$  полагаются известными. Если рассматривать систему, составленную из уравнений 1, 2, 4, 5 и т.д., то неизвестными будут только коэффициенты матрицы вращения  $\mathbf{R}$  и вектора переноса  $\mathbf{r}_0$ , а также величина  $z^{(0)}$ . Эта система, к сожалению, нелинейна, однако, может быть решена приближенными методами. При этом, взяв несколько последовательных кадров и проследив на них одну точку, можно определить параметры движения твердого тела.

Таким образом, при определении параметров движения твердого тела также будет оцениваться и его трехмерная форма. В этом смысле собственное движение объектов оказывается ценным источником информации даже при статичной камере. Интересно, что некоторые животные, например, совы и богомолы, осуществляют движения головой для определения трехмерной формы статичных объектов.

Задача определения параметров движения камеры по наблюдаемому видеоряду называется *пассивной навигацией*. Важными частными случаями является движение камеры вдоль или поперек оптической оси и вращение камеры вокруг какой-либо оси без перемещения. По сути, эти случаи уже обсуждались при рассмотрении проблемы сопоставления изображений трехмерных сцен, поэтому здесь мы к ним не возвращаемся. Напомним лишь, что при движении камеры ее траектория может быть восстановлена с точностью до масштабного множителя. При этом в задаче пассивной навигации в дополнение к межкадровому отождествлению сопряженных точек возникают проблемы отслеживания точек на всех кадрах видеоряда и согласования результатов измерения дальности по разным парам соседних кадров.

### Вопросы и упражнения

1. Как определяются оптический поток и поле движения? В чем между ними различие?
2. Какими будут оптический поток и поле движения в случае однородной вращающейся или неподвижной сферы, освещенной движущимся или неподвижным источником освещения?
3. Какой вид имеет уравнение оптического потока?
4. Какие трудности возникают при использовании оптического потока?
5. На какое направление проекция вектора скорости определяется, исходя из уравнения оптического потока? Каким образом может быть

- восстановлена информация о скорости в перпендикулярном направлении?
6. В каких точках изображения оптический поток определен быть не может?
  7. Как осуществляется доопределение оптического потока?
  8. Какие ограничения на вид оптического потока накладывает предположение о движении твердого тела?

### 4.3. Понимание изображений

Системы компьютерного зрения часто выступают в качестве компонента системы управления или принятия решения, предназначенной для решения некоторой задачи в реальном мире, которая обычно ставится в терминах объектов и ситуаций. Восстановление трехмерных форм и движения объектов, описанных в терминах совокупности поверхностей, оказывается недостаточным.

Представим, к примеру, систему управления транспортным средством. Информация о трехмерной структуре сцены является весьма полезной, так как позволяет устанавливать направление возможного движения и обнаруживать препятствия. Однако по трехмерным характеристикам обочина или тротуар могут мало отличаться от проезжей части. Также в качестве препятствий не будут различаться пешеход и куст на разделительной полосе. Для принятия правильных решений, особенно, в аварийно опасных ситуациях крайне важно такое различие устанавливать.

Таким образом, от систем компьютерного зрения может требоваться не только восстановление физических характеристик объектов, но и построение осмысленного описания сцен. Семантическая интерпретация изображений называется *пониманием изображений*.

Проблемы понимания изображений могут варьироваться от распознавания или обнаружения единичных объектов по их изображениям до построения полного описания сцены в рамках некоторой системы представления знаний. Иногда под системами понимания изображений подразумевают системы обработки изображений, включенные в системы искусственного интеллекта, однако такое определение (даже если его и считать верным) не говорит о содержательной стороне проблемы понимания изображений.

Следует также отметить, что в контексте проблемы понимания изображений термин «понимание» обычно используется в смысле *процедурной семантики*: считается, что система «понимает» смысл изображений или языковых высказываний, если может правильно действовать в соответствии с ними. К примеру, если система может выполнить команду «принеси зеленую пирамиду, расположенную на красном кубике», то считается, что система понимает смысл этого высказывания, равно как понимает и содержание изображений, на основе которых выполняет команду.

### Распознавание объектов

Часто для «понимания» содержания изображения в рамках решаемой задачи системе компьютерного зрения может быть достаточно детектировать или распознавать отдельные объекты. К примеру, если ее рассматривать в качестве компоненты системы экологического мониторинга, то требуется обнаруживать возникновение пожаров, наводнений или других стихийных бедствий без составления полного описания сцены (рис. 4.13).



Рис. 4.13. Пример обнаружения дыма на изображении, что можно трактовать как простейшего «понимание» системой компьютерного зрения содержания изображения в рамках решаемой задачи

Распознавание объектов по их изображениям опирается на методы сопоставления изображений и может также выполняться в рамках различных представлений изображений. Однако при распознавании производится сравнение текущего изображения с несколькими эталонными изображениями (вернее, с результатом их отображения в выбранное представление). Таким образом, одним из ключевых моментов распознавания оказывается определение меры сходства изображений, на основе которой можно осуществить отнесение изображения к одному из классов.

В качестве такой меры сходства достаточно часто используется та же мера сходства, которая привлекается в процессе сопоставления пары изображений. Однако к мере сходства в задачах распознавания предъявляются более жесткие требования. К примеру, величина корреляции двух изображений

$$C_{f_1, f_2}(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1(x, y) f_2(x - \Delta x, y - \Delta y),$$

на основе которой может выполняться сопоставление изображений, не подходит для задачи распознавания. Действительно, простое повышение яркости одного из изображений приведет к повышению величины корреляции, хотя сходство изображений при этом не увеличится (повышением яркости можно добиться большого значения корреляции даже для несходных изображений). Более подходящим при распознавании будет коэффициент корреляции, который нормирован на среднеквадратичную яркость обоих изображений.

При чамферном сопоставлении меру Хаусдорфа тоже оказывается необходимым нормировать, поскольку на ее значение оказывают влияние не

только сходство и различие двух изображений, но также и характеристики каждого из изображений в отдельности. Например, для небольшого контура на большом изображении легко найти фрагмент, на котором мера Хаусдорфа будет равна нулю, в то время как наложить два больших изображения с достижением малого значения этой меры значительно сложнее.

В случае структурного сопоставления мера сходства двух изображений, как правило, основывается на числе сопоставленных структурных элементов и среднеквадратичном отклонении их положений от глобального пространственного преобразования. Однако, опять же, в задачах распознавания этого может быть недостаточно.

Наиболее разработанными методами распознавания являются дискриминантные методы распознавания образов, работающие в рамках признаковых представлений. В этих методах каждый класс объектов отделяется от других совокупностью поверхностей в пространстве признаков, либо строятся статистические модели, на основе которых можно установить вероятность принадлежности любого объекта тому или иному классу. К сожалению, описание изображений в терминах векторов признаков сопряжено с потерей большого объема информации, что существенно сужает сферу применения этих методов.

Наиболее существенной проблема выбора адекватной меры сходства оказывается в задачах извлечения изображений из баз данных. В этих задачах количество изображений, среди которых требуется выбрать наиболее приемлемое, является чрезвычайно большим, но, к счастью, при этом в качестве решения может выступать не одно, а множество изображений. Тем не менее, на настоящий момент для решения этой проблемы привлекают, как правило, текстовые аннотации к изображениям.

### *Интерпретация сцен*

Для интерпретации сцен требуется осуществлять выделение на них объектов с последующим распознаванием. Большинство естественных сцен содержит множество самых разнообразных объектов, поэтому распознавание всех объектов на сцене на современном уровне развития систем компьютерного зрения вряд ли возможно и во многих прикладных задачах не нужно.

Обычно для решения задачи достаточно ограничиться распознаванием лишь значимых элементов сцены. К примеру, для системы управления транспортным средством описание сцены может включать выделенные на изображении области, соответствующие дороге, обочине, небу, пешеходам, машинам и прочим препятствиям (без их детального распознавания). Для бытового робота первоочередную значимость может представлять выделение на изображении областей пола, потолка, стен, дверей, статических и подвижных объектов (не разделяемых на классы). При интерпретации аэрокосмических снимков семантические метки должны назначаться областям леса, водоемов, полей, дорог и других антропогенных объектов (возможно, с разделением на несколько крупных классов).

Интерпретация сцен отличается от распознавания совокупности объектов не только необходимостью выделения этих объектов на изображении, но и необходимостью учета допустимого взаимного расположения объектов на сцене. Действительно, вряд ли небо на изображении может располагаться между дорогой и обочиной или пол – между дверью и потолком. Подобные отношения часто описываются в рамках некоторой системы представления знаний.

Рассмотрим пример такого описания в рамках логики предикатов. На рис. 4.14 приведено изображение. В качестве одного из первых этапов интерпретации обычно выступает сегментация изображения на области (рис. 4.15). Конечным результатом интерпретации сцены является назначение семантических меток каждой из областей.



Рис. 4.14. Пример интерпретируемого изображения дороги

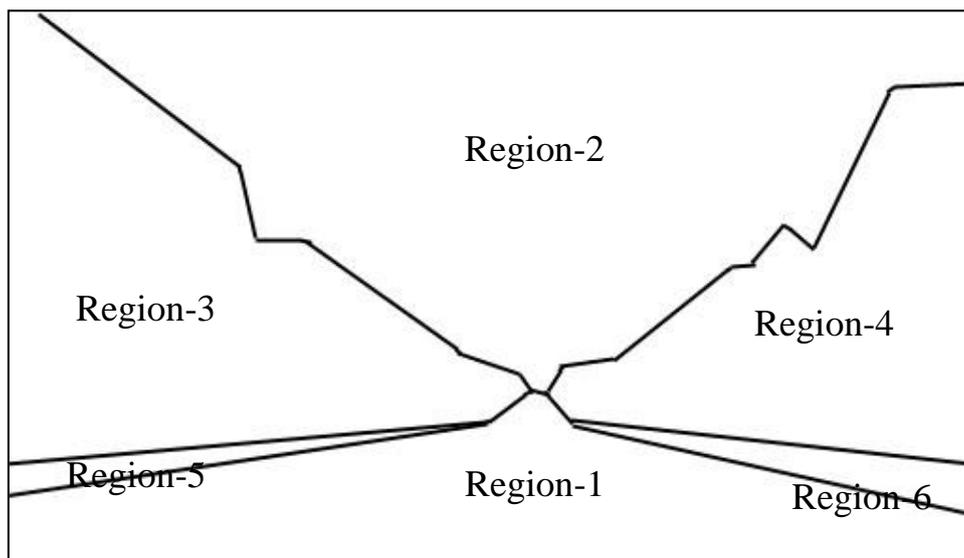


Рис. 4.15. Результат сегментации изображения 4.14 на области

Области на изображении могут быть описаны совокупностью признаков (например, текстурных), и на основе этих признаков каждой области могут быть назначены вероятность ее соответствия одному из допустимых классов.

В данном случае представленные результаты сегментации являются весьма точными, однако на практике они обычно более фрагментарны, что усложняет задачу назначения меток. Привлечение знаний о предметной области может помочь снять неоднозначности интерпретации.

В представленном случае фрагмент базы знаний (общих правил) на языке предикатов может принять следующий вид.

```
region(X) & grass(X) => green(X)
region(X) & sky(X) & time(Y) => blue(X) & day(Y) V dark(X) & night(Y)
region(X) & region(Y) & grass(X) & sky(Y) => above(Y, X) & below(X, Y)
region(X) & region(Y) & sky(X) & sun(Y) => inside(Y, X)
```

В результате логического вывода должны быть получены частные факты

```
forest(Region-3) & forest(Region-4)
road(Region-1)
ground(Region-5) & ground(Region-6)
top(Region-2)
above(Region-3, Region-5)
to-left(Region-3, Region-4)
```

Здесь в качестве переменных могут выступать как элементы изображения (области, структурные элементы, значения каких-либо признаков), так и неизобразительные данные (время суток и года, текущие географические координаты, скорость движения и т.д.). Для некоторых из них могут быть заданы частные факты, полученные в результате обработки изображения (цвет той или иной области и т.д.) или из внешних источников (например, текущее время).

Иногда все элементы системы компьютерного зрения представляют в виде компонентов некоторой системы представления знаний. При этом выделяют три типа знаний: *перцепционное*, *семантическое* и *функциональное* знание. Перцепционное знание позволяет интерпретировать изображения в терминах линий, областей и т.д. (по сути, перцепционное знание реализуется в форме процедур анализа изображений). Семантическое знание описывает определенные абстрактные понятия, такие как форма или конкретные объекты и отношения между ними (приведенный выше пример относится именно к семантическому знанию). Функциональное знание предназначено для регулирования процесса интерпретации изображения в зависимости от предметной области и поставленной цели, то есть включает подсистему манипулирования знаниями (например, логического вывода).

Существуют следующие подходы к представлению знаний в системах интерпретации изображений: 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.

Рассуждения в системах интерпретации сцен сопряжены с существенной долей априорной неопределенности, поэтому классические процедуры логического вывода здесь используются нечасто. Вместо этого может быть использован какой-либо вариант байесовских рассуждений или нечеткий вывод с постепенным выбором наиболее согласованной интерпретации (рис. 4.15).

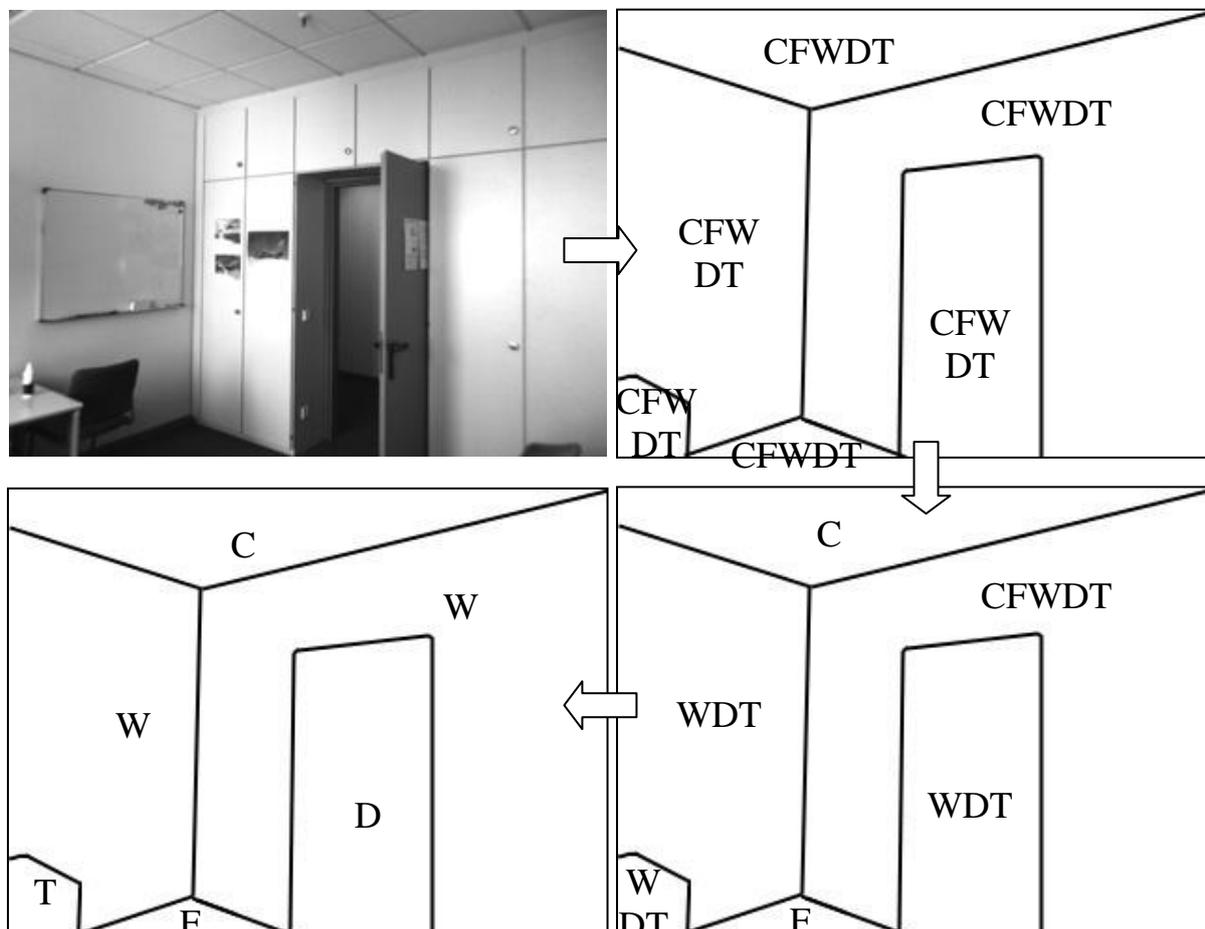


Рис. 4.15. Пример последовательного устранения неопределенности интерпретации на основе знаний

Создание баз знаний является крайне трудозатратной операцией. Еще более сложной проблемой оказывается создание баз знаний, в которые интегрированы системы компьютерного зрения, поскольку оказывается необходимым в явном виде указывать взаимосвязи между перцептивным и семантическим (а также функциональным) знанием, чего не требовалось в классических экспертных системах. Не удивительно, что системы понимания изображений, основанные на знаниях, создаются, как правило, крайне узкоспециализированными.

Как и в случае классических экспертных систем, ключевой проблемой здесь оказывается автоматическое приобретение знаний. Таким образом, для создания систем понимания изображений общего назначения необходимо решение проблемы машинного обучения на основе перцептивного опыта. Существующие подходы в данной области заключаются в имитации процесса усвоения языка младенцем, который исходно не обладает никакими

знаниями о свойствах различных объектов, однако способен научиться строить словесные описания сцен. Рассмотрим в качестве одного из типичных примеров данного подхода серию работ, выполняемых в Media Laboratory Массачусетского Технологического Института.

Развитие упомянутой серии работ началось с системы, получившей название CELL (Cross-channel Early Lexical Learning), которая впоследствии той же группой исследователей была развита в системы Describer, Newt, Ripley, Fuse. Далее мы кратко рассмотрим базовые принципы развиваемого в перечисленных выше работах подхода на примере системы CELL.

### *Общая архитектура системы CELL*

Представим себе машинную систему, оснащенную сенсорами, возможно различных модальностей. В описываемом подходе предполагается, что информация, поступающая от сенсоров, разделяется на каналы, являющиеся либо лингвистическими, либо семантическими (контекстными). Введение информационных каналов в дополнение к сенсорным модальностям вызвано тем, что как лингвистическая, так и семантическая информация может поступать от сенсоров любой модальности.

К примеру, лингвистический канал может содержать извлеченную из акустического сигнала речь или извлеченные из визуального сигнала жесты, движение губ или письменную речь. Тактильный сигнал также может нести для человека лингвистическую информацию. В то же время, любая из этих сенсорных модальностей может нести и семантическую информацию. Например, семантический канал может содержать описание формы, извлеченное из визуального или тактильного сигнала, описание цвета, движения и т.д.

В системе CELL в соответствующих каналах лингвистическая или семантическая информация кодируется посредством системы признаков, процедура выделения которых считается данной априорно. Для поступающих по каждому каналу векторов признаков можно решить задачу их группирования и сформировать классы образов. Если речь идет о семантических признаках, то подобные классы соответствуют некоторым *семантическим категориям*, а если о лингвистических – *лингвистическим единицам*. В системе CELL в процессе обучения строятся классы образов, описываемые эталонным образом, прототипом, и величиной допустимых отклонений от прототипа, при которых образ все еще относится к данному классу.

Лексемой в данной работе называется концепт, включающий в себя некую лингвистическую единицу, ассоциированную с некоторой семантической категорией. Сложность заключается в том, что требуется не просто связать лингвистическую единицу и семантическую категорию, но одновременно с этим и построить соответствующие им классы образов, поскольку они не даются системе априорно.

Формирование классов образов для каждого из каналов в отдельности можно было бы осуществлять в стиле обучения без учителя. Однако

богатство поступающей информации столь велико, что надежно выделять классы затруднительно. Таким образом, следует решать задачи построения классов и их связывание одновременно. Одна из основных идей заключается в том, что использование лингвистической информации позволит надежнее выделять релевантные семантические категории, а использование семантической информации (контекста) – надежнее вести обработку лингвистической информации (выделять и осуществлять классификацию отдельных лингвистических единиц из потока слитной речи). Посмотрим, как это было сделано в описываемой работе.

В ней для семантического и лингвистического каналов вводится понятие S-, и L-событий. Событие – это последовательность выделенных признаков в интервале времени, в котором наблюдается активность в соответствующем канале. Для лингвистического канала это соответствует цепочке некоторых лингвистических признаков, возможно разделенных паузами, соответствующих обособленному речевому высказыванию. Например, при устном вводе L-событие соответствует цепочке произнесенных вслух слов и может быть обнаружено детектором речь/молчание. S-событие соответствует заметным изменениям в сенсорном канале, например, достаточно быстрому движению в поле зрения камеры. Критерий значимости изменений в канале не вводится и предлагается определять его эвристически для каждого отдельного канала.

Итак, пусть в каждом канале выделены события как цепочки признаков. Как L-, так и S- событие может включать образы, относящиеся к различным классам. Такие протяженные во времени события необходимо разбить на некоторые более короткие временные интервалы, сегменты. Для лингвистических событий эти временные интервалы определяют гипотетические границы лингвистических единиц. Если используется несколько лингвистических каналов (например, фонемный и движение губ), то гипотезы об интервалах могут взаимно проверяться и делаться более достоверными. В семантических каналах также необходимо ввести разделение на сегменты. Границы сегментов соответствуют резкому (во времени) изменению скорости, контраста, цвета и т.д. Это означало бы изменение распознаваемых семантических категорий, если бы они были уже сформированы. В связи с этим вводится понятие L- и S-подсобытий, как подпоследовательностей в цепочках признаков, составляющих событиям.

Речь, обращенная к младенцам, обычно относится к мгновенному контексту. Значит, при овладении лексикой является принципиальным одновременное появление лингвистических единиц и соответствующих им семантических категорий, представленных в каналах L- и S-подсобытиями. В описываемой работе моделируется восприятие младенцев и используются записи игры с ними, поэтому делается предположение, что L- и S-подсобытия, относящиеся к одному и тому же объекту, присутствуют в перекрывающихся во времени L- и S-событиях. Когда такие перекрывающиеся события встречаются, то формируется единое LS-событие, которое и подлежит дальнейшему анализу. Это событие помещается в

кратковременную память, являющуюся FIFO-буфером (очередью). Размер этого буфера в системе CELL был взят  $7 \pm 2$  LS-событий. Помимо ссылки на моделирование человеческого восприятия, другим аргументом является экономия вычислительных ресурсов, поскольку поиск в кратковременной памяти повторяющихся событий является исчерпывающим и требует экспоненциального времени.

После помещения LS-событий в буфер требуется выделить подсобытия, в частности, необходимо осуществить сегментацию слитной речи на отдельные слова при неизвестном лексиконе. При этом используется предположение, что в помещенных в буфер LS-событиях значимые подсобытия неоднократно повторяются, поэтому их можно выделить как наиболее длинные повторяющиеся (с определенной точностью) цепочки признаков. Идея поиска повторяющихся подсобытий основывается на том, что речь, обращенная к младенцам, является чрезвычайно избыточной. Если младенцу дают в руки какую-то игрушку, то ее название повторяется много раз («посмотри, какой мячик; мячик круглый, ...»). Поскольку подсобытия повторяются неточно, то чтобы их искать, вводится (в качестве дополнительной априорной информации) метрика в каждом L- и S-каналах. Пары повторяющихся и совместно встречающихся L- и S-подсобытий помещаются в память следующего уровня, имеющую значительно больший объем. Если во время обнаружения следующих LS-событий появляются похожие (в смысле близости в пространстве признаков) значимые пары подсобытий, то на их основе строится лексема (концепт), состоящая из эталонной семантической категории, эталонной лингвистической единицы и радиусов в пространствах признаков, определяющих границы концепта. Это аналогично инкрементному обучению без учителя.

Принятие решения о формировании лексемы осуществляется на основе взаимной информации, содержащейся в соответствующих L- и S-подсобытиях. Пусть имеется набор пар совместно встречающихся подсобытий  $(L_i, S_i)$ . Тогда для каждой из них можно оценить безусловные вероятности  $P(L)$  и  $P(S)$  их появления, а также вероятность совместного появления  $P(L, S)$ . Большое значение взаимной информации, вычисляемой как  $\log_2 P(L, S) - [\log_2 P(L) + \log_2 P(S)]$ , будет служить индикатором того, что появление одного события может быть предсказано при появлении другого подсобытия.

Если оцененная взаимная информация превосходит некоторый порог, то концепт формируется. В процессе функционирования системы при формировании ею концептов рассматриваются в промежуточной памяти только такие события, которые не распознаются как сформированные лексемы. Мы не будем детальнее разбирать вопросы выделения подсобытий, формирования концептов и т.д., а кратко опишем реализацию системы на основе зрительной и акустической модальностей.

### *Реализация зрительной и акустической подсистем в системе CELL*

Один из возможных вариантов реализации описываемой системы (CELL) использует аудио-видео вход. В этой реализации аудиоинформация получается с микрофона, на который записывается естественная слитная речь воспитателя, обращенная к младенцу. В качестве семантических каналов используются каналы, содержащие информацию о форме и цвете объекта, присутствующего в поле зрения камеры. Ставятся задачи выявления лингвистических единиц и семантических категорий формы и цвета и их связывания в концепт таким образом, чтобы в результате обучения система могла называть предъявленные ей объекты, а также, наоборот, по названию объекта выбирать его из заданного набора.

Для выделения лингвистических и семантических признаков необходимы модально-специфичные процедуры, которые считаются заданными априорно. Посмотрим, как здесь ведется выделение семантических признаков в двух каналах: формы и цвета.

Трехмерная форма объекта в явном виде не восстанавливается. Вместо этого используются сгруппированные описания плоских форм, которые получаются в результате съемки объекта с разных ракурсов. На каждом таком снимке осуществляется разделение на объект и фон, при котором используется предположение о том, что фон является сравнительно однородным (однотонная скатерть, простыня и т.д.), на который помещен крупный предмет близко к центру кадра. Все пиксели, которые не относятся к фону, помечаются как принадлежащие к объектам. Находятся связные области из таких пикселей, и из этих областей выделяется наибольшая расположенная ближе к центру кадра. Также рассматривается случай, когда подходящий объект отсутствует (установлением порога на размер и положение) или не полностью попадает в кадр. Далее анализируется форма и цвет выделенной таким образом области, предположительно соответствующей изображению объекта.

Для представления цвета используется двухцветная гистограмма  $8 \times 8$  ячеек нормализованных цветов ( $r, g$ ), инвариантная к смене интенсивности освещения. Поскольку цвета каждого пикселя нормированы на его уровень яркости, то третий цвет дополнительной информации не несет и в представлении не используется.

Описание формы тоже осуществляется с помощью двухмерной гистограммы. Для ее построения находятся точки, лежащие на границе выделенной области. Для них определяются наклоны касательных, проведенных к границе. Также для каждой пары точек границы определяется нормализованное расстояние. По этим двум параметрам строится гистограмма, которая и описывает форму области. Это представление инвариантно к масштабу и повороту объекта в плоскости изображения.

Как видно, представление изображений весьма простое и опирается на сильные предположения. В нем не выделяется значимая информация, а часть ее теряется. Это связано с тем, что в описываемой работе основное внимание

уделялось проблеме анализа речи (которую мы здесь не рассматриваем). Кроме того, в последующих работах анализ изображений был усложнен.

Система CELL, оснащенная акустическими и видео сенсорами, обучалась по записям игры взрослых с младенцами примерно десятимесячного возраста. Для обучения использовались записи игры разных взрослых с младенцами с привлечением различных объектов (игрушки, мячики, обувь и т.д.). Одна запись относилась к одному предмету. Результаты исследований показали относительную успешность данного подхода. Система научилась по естественным записям распознавать 70% предъявляемых объектов по их названиям, представляемым в акустической форме. Для сравнения, поиск повторяющихся фрагментов в записях речи (без использования видео) позволил выделить только 30% релевантных слов (кроме того, не связанных с какими-либо семантическими категориями).

#### *Дальнейшее развитие системы CELL*

Система CELL является наиболее ранней и имеет ряд расширений: это системы Describer и Newt, робот Ripley, система Fuse и ряд других.

В системах Describer и Newt семантическая основа подводится не только под отдельные слова, но и под целые выражения, описывающих пространственные отношения между объектами, такие как «большой синий квадрат справа от красного прямоугольника». Обучение системы Describer производится по примерам смоделированных сцен, сопровождающихся выполненными людьми описаниями, по которым можно выбрать нужный объект. Система Newt учится по изображениям реальных сцен, сопровождающихся устным описанием того объекта, на который система указывает лазерной указкой (на сцене может присутствовать несколько объектов, помещенных на однородный фон).

Системы Describer и Newt отличаются в том, что в первой осуществляется описание выбранного объекта на сцене, в то время как во второй – выбор объекта по его словесному описанию. Помимо проблемы связывания слов с семантическими категориями здесь возникают также две дополнительные проблемы: разделение слов на классы (например, слова «желтый», «розовый», «зеленый» должны принадлежать одному классу, а «над», «слева», «рядом» – к другому классу) и определение порядка слов в предложении (обучение синтаксису).

В системе Fuse основное внимание уделяется вопросам взаимодействия зрительной и акустической подсистем на нижнем уровне, а именно, влиянию речи на управление вниманием зрительной подсистемы и влиянию визуальной информации на улучшение сегментации и распознавания слов и понимания речи.

В работе Ripley подход получил дальнейшее развитие по нескольким направлениям. Во-первых, расширена сенсорная система робота за счет добавления в нее проприоцепторной подсистемы, чувства гравитации и тактильной подсистемы (чувства осязания). Во-вторых, добавлена система эффекторов: у робота появилась возможность совершать движения и брать

предметы манипулятором. И, в-третьих, добавлена «ментальная» модель окружающего физического мира, в которой описываются пространственные положения обнаруженных роботом объектов. Расширение системы сенсоров и эффекторов дает роботу возможность в процессе обучения устанавливать семантическую опору для таких слов, как «легкий», «тяжелый», «дотрагиваться», «поднимать», «давать» и т.д. Ментальная модель физического пространства не только позволяет осуществлять в нем навигацию, но и позволяет понимать различия в таких фразах, как «мяч слева от меня» или «мяч слева от тебя». Для этого в ментальной модели реализована возможность смены точки зрения, что осуществляется с помощью библиотеки 3D-моделирования OpenGL, на основе которой и строится воображение робота. Таким образом, слова «мой» или «твой» находят свою семантическую опору в ментальной операции переноса точки зрения.

В результате обучения упомянутые системы способны строить описания сцен и выполнять в них указанные человеком действия, что свидетельствует об определенном уровне «понимания» как изображений, так и речевых высказываний.

### **Вопросы и упражнения**

1. В чем состоит проблема понимания изображений?
2. В чем сходство и различие задач сопоставления и распознавания изображений?
3. Можно ли свести задачу интерпретации сцен к распознаванию объектов?
4. Какие представления знаний используются в системах понимания изображений?
5. Чем отличается семантическое, перцептивное и функциональное знание?
6. Какая основная проблема возникает в системах понимания изображений, основанных на знаниях?
7. Каковы основные идеи, заложенные в систему CELL?
8. Какую семантическую основу в зрительном восприятии имеют существительные, прилагательные, местоимения?

## Список рекомендуемой литературы

### Базовый учебник:

1. Хорн Б.К.П. Зрение роботов / Б.К.П. Хорн. – Пер. с англ. – М.: Мир, 1989. – 489 с.

### Базовое учебно-методическое пособие:

2. Шапиро Л. Компьютерное зрение: Учеб. пособие для вузов / Л. Шапиро, Дж. Стокман. – Пер. с англ. – БИНОМ. Лаборатория знаний, 2006. – 752 стр.

### Основная литература:

3. Форсайт Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – Пер. с англ. – М.: Издательский дом "Вильямс", 2004. – 926 с.
4. Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов / Д. Марр. – Пер. с англ. – М.: Радио и связь, 1987. – 400 с.
5. Потапов, А.С. Распознавание образов и машинное восприятие: общий подход на основе принципа минимальной длины описания / А.С. Потапов. – СПб.: Политехника, 2007. – 548 с.
6. Дуда, Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт. – М.: Мир, 1976. – 511 с.
7. Ту, Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М.: Мир, 1978. – 412 с.
8. Фурман, Я.А. Введение в контурный анализ и его приложения к обработке изображений и сигналов / Я.А. Фурман и др. – М.: ФИЗМАТЛИТ, 2002. – 592 с.
9. Козлов, Ю.М. Адаптация и обучение в робототехнике / Ю.М. Козлов. – М.: Наука. Гл. ред. физ.-мат. лит., 1990. – 248 с.
10. Ерош, И.Л. Адаптивные робототехнические системы / И.Л. Ерош, М.Б. Игнатъев, Э.С. Москалев. – Л.: ЛИАП, 1985. – 144 с.
11. Прэтт, У. Цифровая обработка изображений. Кн. 2 / У. Прэтт. – М.: Мир, 1982. – 480 с.
12. Фу, К. Структурные методы в распознавании образов / К. Фу. – Пер. с англ. – М.: Мир, 1977. – 320 с.
13. Анисимов, Б.В. Распознавание и цифровая обработка изображений: Учеб. пособие для студентов вузов / Б.В. Анисимов, В.Д. Курганов, В.К. Злобин. – М.: Высш. шк., 1983. – 295 с.

### Дополнительная литература:

14. Минский, М. Фреймы для представления знаний / М. Минский. – Пер. с англ. – М.: Энергия, 1979. – 151 с.
15. Искусственный интеллект. Справочник в 3-х томах. Кн. 1. / М.: Радио и связь, 1990. – 461 с.
16. Юревич, Е.И. Основы робототехники. 2-е изд. / Е.И. Юревич – СПб.: БХВ-Петербург, 2007. – 416 с.
17. Хьюбел, Д. Глаз. Мозг. Зрение. / Д. Хьюбел. – М.: Мир, 1990. – 240 с.

**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 КОМПЬЮТЕРНОЙ ФОТОНИКИ И ВИДЕОИНФОРМАТИКИ**

Достижения в оптической науке, технике и технологиях за последние годы способствовали появлению нового направления – фотоники. Этот термин охватывает область науки и техники, связанную с использованием светового излучения (или потока фотонов) в оптических элементах, устройствах и системах.

На рубеже XX – XXI веков электронные информационные технологии достигли фундаментальных и технических пределов производительности при продолжающемся росте потребительского спроса на скорость и объем обрабатываемой и передаваемой информации. Решение данной проблемы потребовало разработки нового поколения информационно – телекоммуникационных систем, основанных на технологиях фотоники. В фотонике появилось новое динамично развивающееся направление, определяющее прогресс мировой науки и техники, – «оптоинформатика». Под «оптоинформатикой» понимают область науки и техники, связанную с исследованием, разработкой, созданием и эксплуатацией новых материалов, технологий, приборов и устройств, направленных на передачу, прием, обработку, хранение и отображение информации.

Изучение фотоники основывается на знании принципов формирования, преобразования, анализа изображений, теории построения информационных систем. Интеграция фотоники и компьютерных технологий позволяет создавать методы, которые возможно реализовать исключительно средствами компьютерной фотоники, обеспечивая развитие технологий качественно нового уровня.

По многим направлениям фотоники и оптоинформатики Россия находится на уровне промышленно – развитых стран (интегральная оптика, системы приема, обработки и отображения информации и др.), а, по некоторым – даже опережает. Приоритетными направлениями являются: волоконная оптика (работы академика Дианова Е.М. – ИОФ РАН), голография (академик Денисюк Ю.Н. – ГОИ им. С.И. Вавилова), полупроводниковые лазеры (академик Алферов Ж.И – ФТИ РАН им. А.Ф. Иоффе), полифункциональные оптические материалы (академик Петровский Г.Т. – ГОИ им. С.И. Вавилова) и др.

Ввиду большого научного и практического значения направления "Фотоника и оптоинформатика", а также спроса на него на потребительском рынке, в 2002 г. в СПбГУ ИТМО был организован факультет «Фотоники и оптоинформатики» под руководством доктора физ.-мат. наук, профессора С.А. Козлова. По инициативе профессорско-преподавательского состава, начиная с 2005 года, на факультете стала работать выпускающая кафедра «Компьютерной фотоники», которую возглавил доктор технических наук, профессор И.П. Гуров.

История кафедры началась в 1946 году. На всех этапах развития результаты научных исследований, проводимых сотрудниками кафедры, неизменно использовались в учебном процессе. Совершенствовались направления подготовки студентов, изменялось название кафедры, но всегда кафедра гордилась своими выпускниками.

Выпускники кафедры занимают видное место в оптической науке: академик РАН Ю.Н. Денисюк, изобретатель трехмерной голографии; член-корр. РАЕН, профессор Н.Г. Бахшиев, известный специалист в области спектроскопии межмолекулярных взаимодействий; Заслуженный деятель науки РФ, профессор Г.Н. Дульнев, крупный ученый в области теплофизики, долгие годы бывший ректором ЛИТМО; профессор И.М. Нагибина, исследования которой в области физической оптики получили широкое признание.

Одной из важнейших задач кафедры является организация учебного процесса и подготовка профессионалов в области компьютерной фотоники. Направление работы кафедры определяется развитием информационных технологий и компьютерных систем в области формирования, синтеза, обработки и анализа изображений на основе интеграции эффективных компьютерных систем с системами фотоники.

Проводимые исследования в области компьютерной обработки когерентных и некогерентных изображений обеспечивают решение научно-технических задач оптической томографии, цифровой голографии, синтеза, анализа, распознавания и классификации изображений.

Научным консультантом работ кафедры в области компьютерной обработки изображений – иконки – является член-корреспондент РАН М.М. Мирошников.

Кафедра проводит работы в рамках международных научных проектов в сотрудничестве с ведущими зарубежными университетами, институтами и исследовательскими лабораториями Италии, Финляндии, Франции, Германии, Великобритании, Японии, США и других стран в области оптической когерентной томографии для биомедицинских исследований, цифровой голографии для исследования микро- и наноструктур, трехмерной фотографии микро- и макроскопических объектов, гиперспектральной обработки изображений.

В последнее время на кафедре активно развивается новое направление – видеоинформатика. Ввиду этого в 2010 году кафедра была переименована в кафедру Компьютерной фотоники и видеоинформатики. В 2011 году

Университет получил статус Национального исследовательского университета, в этом есть и заслуга преподавательского коллектива кафедры Компьютерной фотоники и видеоинформатики. Эти обстоятельства позволяют обеспечивать и в дальнейшем подготовку высококлассных востребованных на рынке специалистов в области Компьютерной фотоники и видеоинформатики.

Потапов Алексей Сергеевич

## **Системы компьютерного зрения**

**Учебное пособие**

В авторской редакции  
Дизайн и верстка

А.С. Потапов  
В.В. Потапова

Редакционно-издательский отдел Университета ИТМО  
Зав. РИО

Н.Ф. Гусарова

Подписано к печати 20.12.2016

Заказ № 3800

Тираж 50 экз.

Отпечатано на ризографе

**Редакционно-издательский отдел**  
**Университета ИТМО**  
197101, Санкт-Петербург, Кронверкский пр., 49