



**Ш. Ульман**

**ПРИНЦИПЫ  
ВОСПРИЯТИЯ  
ПОДВИЖНЫХ  
ОБЪЕКТОВ**

**Ш. Ульман**

**ПРИНЦИПЫ  
ВОСПРИЯТИЯ  
ПОДВИЖНЫХ  
ОБЪЕКТОВ**

Перевод с английского  
Р. М. Абдусаматова и Д. А. Белова  
под редакцией В. Л. Стефанюка



**МОСКВА  
«РАДИО И СВЯЗЬ»  
1983**

УДК 62.506.222.001.57

Ульман Ш. Принципы восприятия подвижных объектов: Пер. с англ. — М.: Радио и связь, 1983. — 168 с., ил.

В книге дается широкий теоретический и экспериментальный анализ проблемы восприятия наблюдаемого человеком и искусственными системами движения объектов в пространстве. Используя тщательно подобранные демонстрационные изображения, автор показывает, что реальный процесс восприятия разбивается на ряд этапов, из которых выделяются этап установления соответствия объектов в смежных кадрах и этап последующей идентификации трехмерной структуры движущихся объектов. Считается, что процесс установления соответствия является автономным процессом нижнего уровня зрительной системы, использующим некоторую меру структурного сходства деталей изображения и определенную процедуру локального взаимодействия. Это подтверждается разработанной авторами вычислительной схемой, которая позволяет объяснить наблюдаемые в экспериментах явления. После сопоставления изображений одного и того же физического объекта следующим шагом является интерпретация обнаруженных двумерных преобразований, которая включает в себя выделение объектов в изменяющейся сцене и восстановление их трехмерной структуры и характера движения. На этапе интерпретации существенно используются естественные (рефлексивные) ограничения, в частности предположение о сохранении формы. Свообразие и новизна результатов определяются сочетанием методов искусственного интеллекта с традиционными методами из области психологии и физиологии зрительного восприятия.

Для научных работников в области машинного зрения, робототехники и искусственного интеллекта.

Табл. 1. Ил. 55. Библиогр. 144.

Редакция переводной литературы

SHIMON ULLMAN

THE INTERPRETATION OF VISUAL MOTION

The MIT Press

Cambridge, Massachusetts, and London, England

У 1502000000—114  
046(01)—83 89—83

© 1979 by The Massachusetts Institute of Technology  
© Перевод на русский язык, предисловие, примечания.  
Издательство «Радио и связь», 1983

## Предисловие редактора перевода

Предлагаемая советскому читателю книга Ш. Ульмана вышла в Массачусетском технологическом институте в США в серии книг по искусственному интеллекту — направлению, которое уже два десятилетия бурно развивается, привлекая все новых сторонников. О последнем свидетельствует большое число разнообразных докладов, представленных на таких важных форумах, как VII Международная объединенная конференция по искусственному интеллекту (Канада, 1981 г.) и Европейская конференция по искусственному интеллекту (Франция, 1982 г.).

Сегодня мало кого смущает отсутствие согласованного определения, что же такое искусственный интеллект, поскольку теоретическая и прикладная значимость результатов, полученных в этой области, огромна. Особенно важна, как стало ясно в последние годы, роль подходов и методов, развиваемых в области искусственного интеллекта, для создания новой вычислительной техники.

С другой стороны, трудно не согласиться с профессором П. Уинстоном в том, что искусственный интеллект — это попытка с помощью вычислительных или «компьютерных» моделей разобраться в том, что такое интеллект вообще. Зрительная система человека дает яркий пример в высшей степени «интеллектуальной» обработки поступающей в органы зрения информации. Однако многие закономерности и принципы такой обработки пока что не ясны. Прогресс в этом направлении послужит сразу двум целям: пониманию процесса зрительного восприятия человека и открытию дополнительных возможностей в искусственных зрительных системах, например в случае промышленных или исследовательских роботов.

Особенностью книги Ш. Ульмана, подчеркнутой ее названием, является обращение к новой для искусственного интеллекта области восприятия подвижных объектов (прежде сцена, состоящая из кубиков, призм, деталей машин и т. п., считалась фиксированной и неподвижной). При этом автор концентрирует внимание на вычислительных аспектах восприятия движения, т. е. на выяснении характера вычислений, осуществляемых при зрительном восприятии, а также механизмах, их обеспечивающих.

Основные вошедшие в книгу результаты получены в Массачусетском технологическом институте в результате тесного контакта факультета электротехники и вычислительной техники (руководи-

тель — профессор М. Иден), в котором работает Ш. Ульман, с лабораторией искусственного интеллекта (руководитель — профессор П. Уинстон). Дело не только в мощной технической базе лаборатории искусственного интеллекта, но и в том, что в результате такого контакта сам подход к анализу работы зрительной системы претерпел значительные изменения по сравнению с традиционными исследованиями, ведущимися под руководством М. Идена.

Кратко это отличие состоит в особом подчеркивании символического характера обработки в ходе восприятия. Делается вывод о двухэтапности процесса зрительного восприятия: на первом этапе выделяются элементарные символы, характеризующие данную зрительную обстановку, а на втором — производится анализ наблюдаемой сцены в терминах этих символов.

Как всякая четко выраженная схема, вряд ли такая модель характеризует все случаи зрительного восприятия, и в этом плане работы Ш. Ульмана подвергались в литературе некоторой критике. Однако достоинствами схемы являются ее прозрачность и практическая ценность: она легко воссоздается искусственно, причем на этапе вторичной зрительной обработки можно воспользоваться богатым арсеналом методов символической обработки, накопленных в области искусственного интеллекта.

В этом кратком предисловии не стоит давать исчерпывающую оценку книги Ш. Ульмана, тем более заниматься ее подробным критическим разбором. Хотелось бы лишь отметить, что она несомненно будет с интересом прочитана представителями многих научных и инженерных специальностей, чему немало способствуют легкий и ясный способ изложения, оригинальность авторского подхода, единая теоретическая концепция, богатый набор простых для понимания и, как правило, убедительных экспериментов.

Перевод первой части книги выполнен Р. М. Абдусаматовым, а второй части — Д. А. Беловым.

*В. Л. Стефанюк*

### **От редакции серии книг по искусственному интеллекту Массачусетского технологического института**

Искусственный интеллект изучает интеллектуальную деятельность с использованием идей и методов, связанных с применением вычислительных машин. К сожалению, в настоящий момент дать определение интеллекта, по-видимому, невозможно, поскольку представляется, что интеллект сочетает в себе самые разнообразные способности к обработке информации и ее представлению.

Разумеется, в психологии, философии, лингвистике и связанных с ними разделах науки можно найти различные методы и подходы к изучению интеллекта. По большей части, однако, пред-

ложенные в этих областях науки теории являются слишком неполными и расплывчатыми, чтобы их можно было реализовать на языке вычислений. Необходимо нечто существенно большее, хотя ценные мысли, важные взаимоотношения и ограничения могут быть заимствованы и из традиционных исследований того, что, в конце концов, является впечатляющим доказательством фактического существования интеллекта.

Искусственный интеллект опирается на новый подход и новую методологию. Главная его цель — сделать вычислительные машины разумными, чтобы, с одной стороны, сделать их более полезными, а с другой — понять принципы, лежащие в основе интеллектуальной деятельности. Тот факт, что «разумные» ЭВМ будут крайне полезны, совершенно очевиден. Более глубоким является соображение о том, что искусственный интеллект нацелен на проникновение в сущность интеллектуальной деятельности, используя идеи и методы из области вычислений и предлагая, тем самым принципиально новую и отличную от предшествующих основу построения теории интеллекта. Большинство исследователей, занятых созданием искусственного интеллекта, полагают, что такого рода теории будут приложимы к любому интеллектуальному устройству переработки информации, биологического происхождения или искусственно созданного.

Кроме того, существуют побочные эффекты, также заслуживающие внимания. Любая программа, которая достаточно хорошо будет моделировать даже некоторую компоненту интеллекта, в силу своей природы неизбежно окажется объемистой и сложной. Следовательно, искусственный интеллект то и дело сталкивается с ограничениями, налагаемыми существующим уровнем развития вычислительной техники. Возникающие при этом трудности в достаточной мере серьезны и интересны, чтобы привлечь исследователей в области искусственного интеллекта к упорной работе по их преодолению. Поэтому естественно, что налицо непрерывный поток идей из области искусственного интеллекта в область вычислительных наук и нет никаких признаков того, что этот поток когда-либо иссякнет.

Цель серии книг по искусственному интеллекту, издаваемой Массачусетским технологическим институтом (МТИ), — обеспечить как специалистов, так и студентов в разных областях науки своевременной и подробной информацией о том, что происходит на переднем крае науки в исследовательских центрах всего мира.

*Патрик Генри Уинстон  
Майк Брэйди*

### **Предисловие автора**

Исследование, описанное в настоящей книге, отражает эволюцию подходов к изучению зрительного восприятия. Когда я заинтересовался явлениями зрительного восприятия впервые, то надеялся понять их через изучение анатомии и физиологии зрительной системы. По причинам, о которых будет сказано во введении к книге, постепенно все яснее становился тот факт, что изучение на более абстрактном уровне вычислений, осуществляемых в ходе зрительного анализа, может существенно дополнить, а где-то и направить изучение работы самой зрительной системы. Используемый в области искусственного интеллекта подход вида «сделай сам» обещал определенные преимущества. Поэтому в основе всей работы лежало стремление воспользоваться искусственными схемами, которые реально могут выполнять некоторые функции, присущие биологической зрительной системе, в качестве средства исследования задач, стоящих перед зрительной системой, и тех методов, которыми эти задачи решаются.

В лаборатории искусственного интеллекта Массачусетского технологического института мне посчастливилось встретить Д. Марра, который в то время закладывал основы вычислительной теории зрения. Я глубоко обязан ему за ту неоценимую поддержку, которую он мне оказал, и за тесное и стимулирующее сотрудничество. Изучение зрения с позиций вычислений, проводимое в МТИ, постепенно привело к созданию неформальной, но интенсивно взаимодействующей группы исследователей из МТИ и других мест. Я много вынес из общения с большинством членов этой группы: Т. Поджио из Института биологической кибернетики им. М. Планка в Тюбингене, У. Ричардсом с факультета психологии МТИ, К. Стевенсом, Е. Гримсоном и Е. Хилдретом из лаборатории искусственного интеллекта МТИ, сделавшими много ценных замечаний к черновикам этой книги, К. Нисихарой и др. Настоящая работа была бы невозможна без неповторимой научной атмосферы и возможностей, представляемых лабораторией искусственного интеллекта.

Я благодарен Институту перспективных вычислений в Саннивейле, Калифорния, за их приглашение и поддержку, оказанную мне летом 1977 г., журналу «Труды Королевского общества» за разрешение воспроизвести части работы [Ульман, 1978b], включая помещенные там рис. 1 и 2, и Карен Прендергаст за подготовку рисунков.

Наконец, я хочу выразить признательность моей жене Чане за ее ободрение и поддержку.

Настоящая книга излагает результаты исследования, проведенного на базе лаборатории искусственного интеллекта Массачусетского технологического института.

## ВВЕДЕНИЕ

### 1. Предварительные теоретические соображения

Интерпретация наблюдаемого движения — это процесс, с помощью которого на основе попадающего в глаза меняющегося изображения создается описание окружающей обстановки в терминах объектов, их трехмерной формы и перемещения в пространстве.

Исследование, описанное в этой книге, направлено на создание вычислительной теории такого процесса. Таким образом, оно преследует двоякую цель: во-первых, исследовать сам процесс интерпретации наблюдаемого движения; во-вторых, проиллюстрировать вычислительный подход к изучению зрительного восприятия. В настоящем введении кратко обсуждаются некоторые общие проблемы, лежащие в основе вычислительного подхода. Другие проблемы обсуждаются в последующих главах (в основном в гл. 3 и 4), причем обсуждение сопровождается примерами из теории наблюдаемого движения.

В основе вычислительной теории зрительного восприятия лежит представление о том, что зрительную систему человека можно рассматривать как систему, оперирующую символами [Марр, 1976]. Обеспечиваемые ею вычислительные процессы сводятся, по крайней мере отчасти, к созданию *практически полезных описаний наблюдаемого окружения*. Непосредственным следствием такой точки зрения является различие между физическим воплощением символов, которыми манипулирует система, с одной стороны, и значениями этих символов — с другой. Иными словами, *вычисления*, производимые системой, можно изучать почти независимо от физических *механизмов*, лежащих в их основе.

По-видимому, до того как появилось само понятие «вычислительная машина», было довольно трудно уяснить ту роль, какую физическое устройство может играть в обработке информации. Одним из последствий этого затруднения явилась концентрация сил исследователей на изучении физической структуры мозга как средстве для понимания его функций. В результате развития вычислительной техники и теории вычислений стало ясно, что можно провести определенную грань между изучением процесса на абстрактном уровне и изучением физической структуры, реализующей этот процесс. Сама физическая система и реализуемое ею вычисление, конечно, связаны между собой, но их не надо отождествлять. Одно из соображений в пользу отделения меха-



низма устройства от вычисления состоит в том, что не все физические явления в системе, манипулирующей символами, являются осмысленными с точки зрения вычислений. Позвольте проиллюстрировать эту мысль на примере простого вычислительного устройства — электронного калькулятора. Некоторые явления в электронном калькуляторе несут определенный смысл в рамках производимых арифметических действий. Другие явления в калькуляторе и некоторые его компоненты, например относящиеся к блоку питания, такого смысла не несут. Следовательно, теория электронного калькулятора и теория производимых им вычислений различны и неизоморфны. Однако различие между механизмом и вычислением идет дальше этой неизоморфности: логика, управляющая вычислением, не может быть полностью выражена в терминах физической системы. Например, тот факт, что обыкновенный карманный калькулятор дает только первые восемь цифр квадратного корня из числа 2, является свойством конкретного устройства. Тот же факт, что это число не может быть представлено каким-либо конечным десятичным числом, относится уже к теории действительных чисел. Кроме того, теория механизма и теория вычисления имеют дело с разными объектами. Теория калькулятора связана с электронными схемами, токами и напряжениями. Теория же того, что вычисляется, имеет дело с арифметическими объектами.

Подведем итог этим рассуждениям. Каждый раз, говоря о том, что определенное устройство можно рассматривать как символическую систему, мы подразумеваем, что происходящие в ней некоторые события можно непротиворечиво истолковать как объекты, наделенные смыслом в рамках некоторой смысловой системы. Таким образом, можно исследовать сам физический прибор или изучать то, что он вычисляет. Эти два способа исследования неидентичны, и ни один из них нельзя свести полностью к другому.

Из проведенных рассуждений применительно к изучению зрительной системы следует, что анализировать следует, по крайней мере, две взаимосвязанные, но различные составные части: вычисления и обеспечивающие их механизмы. Исследование с вычислительной точки зрения относится к природе внутренних представлений, используемых зрительной системой, и к природе порождающих их процессов. Исследование механизмов касается (в случае биологических зрительных систем) нейронных систем, в которых эти механизмы реализованы [Марр и Поджио, 1977].

Рассмотренная дихотомия между механизмом и вычислением приводит к возникновению следующей проблемы: «Как определить, какие вычисления производит зрительная система?». Один из возможных подходов к этой проблеме состоит в том, чтобы использовать результаты анатомических и физиологических исследований реальной физической системы для выявления производимых вычислений. Другой подход, распространенный в психологических исследованиях, состоит в изучении наблюдаемого поведения системы на языке определенных соответствий между входом

выходом. В данной работе акцент делается на третьем подходе — на изучении задач, решаемых зрительной системой, с целью выяснения выполняемых ею вычислений. Такое исследование начинается с определения конечной цели решения имеющейся зрительной задачи. При точной формулировке этой цели ключевые значения могут иметь психофизические данные. Например, исходя из исследований Юлеша по стереоскопическому зрению [Юлеш, 1971] можно сказать, что одна из целей вычислений может быть сформулирована так: выделение на паре изображений карты испаратности без использования монокулярных признаков. Как только сформулирована цель, можно обратиться к проблемам, возникающим при ее достижении. Цель этих усилий — сформулировать теорию, объясняющую соответствия между входом и выходом системы при идеализированных условиях<sup>1\*</sup>. При формулировке вычислительной теории многое зависит от выявления неявных допущений, используемых зрительной системой. Если говорить кратко, речь идет о важных предположениях относительно окружающей обстановки, учитываемых в процессе вычисления. Они будут подробно исследованы в гл. 4, в которой обсуждаются понятие рефлексивных ограничений<sup>\*\*</sup>.

За созданием вычислительной теории может следовать изучение конкретного алгоритма, используемого зрительной системой при достижении рассматриваемой цели. Одной из побочных причин создания подобного алгоритма является то, что он помогает привлечь внимание на проблемы, которые иначе могли бы остаться незамеченными. Различие между вычислительной теорией и реализующим ее конкретным алгоритмом до некоторой степени аналогично различию между математической функцией и программной процедурой, используемой в ЭВМ. Из них первая обладает более общей природой и определяет соответствия между входом и выходом в целом, тогда как вторая определяет конкретную реализацию с применением данного множества элементарных операций (обсуждение методов и алгоритмов см. в работе [Марр, 1977a]). При ограничении множества подлежащих рассмотрению алгоритмов полезными могут оказаться психофизические данные. Это множество можно дополнительно сузить, потребовав, чтобы алгоритм был осуществим в рамках биологической системы (см. гл. 3).

В этой работе мы подходим к исследованию проблемы интерпретации наблюдаемого движения, подразделяя ее на проблему *установления соответствия* и проблему *трехмерной интерпретации*.

\* Арабские цифры указывают на авторские примечания, помещенные в конце книги. — *Прим. пер.*

\*\* Большой опыт построения разнообразных искусственных систем для достаточно сложных предметных областей показал, что их создание было бы невозможно без явного учета ограничений, налагаемых особенностями данной предметной области. В противном случае возникающие вычислительные задачи приводят к громоздкому перебору, выходящему за рамки возможностей современных ЭВМ. Автор считает, что подобная ситуация справедлива и для реальных биологических систем. — *Прим. ред.*

Первая состоит в идентификации некоторой части изменяющегося зрительного массива как некоторого одиночного объекта, находящегося в состоянии движения или изменения. Само понятие «установление соответствия» возникает тогда, когда зрительная задача рассматривается (как это и будет в значительной части настоящей работы) при наличии последовательности изображений, таких как кадры кинофильма. В этих условиях возникает проблема установления соответствия между частями предыдущего и последующего кадров, представляющих тот же самый объект в последующий момент времени.

После того как будут идентифицированы различные проекции одного и того же объекта, следующим шагом явится интерпретация изменения в наблюдаемом изображении объекта в терминах движения или видоизменений, свойственных самому объекту. Проблема установления соответствия рассмотрена в первой части данной работы (см. гл. 1—3), тогда как вторая часть (см. гл. 4 и 5) посвящена исследованию процесса трехмерной интерпретации.

## 2. Методологические соображения

1. *Дискретное представление.* В большей части настоящей работы проблема установления соответствия рассматривается в условиях дискретного процесса стимуляции, а именно стимуляции в виде последовательности кадров типа кинематографических. Воспринимаемое в таких условиях движение носит название «кажущегося движения».

Кажущееся движение. Если два стимула, например две световые полоски, предъявляются последовательно с соответствующим временным интервалом, то наблюдается явное движение, направленное от первого стимула ко второму. Такое движение называют также оптимальным или бета-движением. Возникновение его зависит от нескольких факторов, из которых наиболее полно изучены временные интервалы, пространственная удаленность стимулов друг от друга и их интенсивность [Вертгеймер, 1912; Корте, 1915; Холвэнд, 1935; Нефф, 1936; Корбэн, 1942; Грэхэм, 1965; Канеман и Вольман, 1970; Эттнев и Блок, 1973]. Фундаментальное отличие действительного движения от кажущегося заключается в свойстве непрерывности предъявляемого стимула, а не в его реальности. При так называемом действительном движении стимул перемещается по сетчатке непрерывно, а при кажущемся движении — дискретно.

Условия возникновения кажущегося движения служат удобной основой для изучения процесса установления соответствия. С одной стороны, задача становится проще, поскольку она сводится к сопоставлению дискретных стимулов. С другой стороны, имеются убедительные доводы в пользу того, что пока речь идет о процессе установления соответствия, между кажущимся и действительным движениями нет фундаментальных отличий<sup>2</sup>. Теория установления соответствия при движении, развитая в гл. 3, равно как

и теория трехмерной интерпретации, развитая в гл. 4, приложена к обоим условиям предъявления стимулов.

2. *Использование демонстрационных экспериментов.* В настоящей работе большинство экспериментов использовалось в основном так же, как языковые предложения в лингвистике. Иными словами, они служили для освещения таких положений, которые не требовали особой точности при разработке техники психологического эксперимента. В некоторых случаях (см. разд. 2.5) были желательны точные измерения, которые производились в соответствии с методикой стандартных экспериментальных процедур.

Предъявление демонстрационных изображений проводилось с помощью дисплея DEC GT 44, управляемого вычислительной машиной DEC PDP-11/40, и дисплея DEC 343, управляемого DEC PDP-6. Оба устройства были подсоединены к системе DEC PDP-10, имеющейся в лаборатории искусственного интеллекта МТИ. Типичный демонстрационный эксперимент состоял в показе двух последовательных кадров. Каждый кадр показывался в течение интервала, называемого *временем предъявления*, с периодом между двумя кадрами, называемым *межстимульным интервалом* (МСИ). Всюду, если не оговорено противное, считается, что время предъявления равно 120 мс, а МСИ составляет 40 мс<sup>3</sup>. Все предъявляемые изображения наблюдались одним глазом в слабо освещенной комнате на расстоянии не менее 100 см<sup>4</sup>. Для проведения экспериментов, описанных в разд. 1.2, использовался шестиканальный стереоскопический тахистоскоп марки «Сайентифик прототайп». Он создавал световой уровень 120 кд/м<sup>2</sup> и был снабжен газоразрядной лампой широкополосного флуоресцентного излучения. Дальнейшие детали, касающиеся экспериментов, описаны в работе [Ульман, 1978а].

## ЧАСТЬ I

# ПРОБЛЕМА УСТАНОВЛЕНИЯ СООТВЕТСТВИЯ

### КРАТКОЕ СОДЕРЖАНИЕ

Установление соответствия — это процесс идентификации элементов на различных изображениях, которые представляли бы один и тот же объект, показанный в разные моменты времени, обеспечивающий тем самым целостность восприятия объектов, находящихся в движении или подвергающихся изменению. В первых трех главах настоящей работы процесс установления соответствия исследуется с применением как экспериментального, так и вычислительного подходов. Центральные результаты гл. 1 и 2 заключаются, во-первых, в том, что процесс установления соответствия является автономным процессом нижнего уровня зрительной системы, а во-вторых, этот процесс производит сопоставление элементарных составляющих изображений на основе некоторой встроенной меры структурного сходства и учета локальных взаимодействий. В гл. 3 предпринято изучение этой проблемы с вычислительной точки зрения и разработана практически применимая схема установления соответствия, согласующаяся с результатами гл. 1 и 2.

В гл. 1 мы обращаемся к такой проблеме: «Какие именно элементы сопоставляются в процессе установления соответствия?». В разд. 1.1 теоретически обосновывается, что операции установления соответствия должна предшествовать первичная организация входного необработанного зрительного массива и выделения на нем неких содержательных единиц. Следовательно, основными элементами, называемыми *опорными символами соответствия*, должны быть такие образования, как фрагменты краев и линий, полосы и пятна. В следующих двух разделах приводятся свидетельства в пользу того, что это те самые элементы, которые сопоставляются человеком в процессе восприятия движения. В разд. 1.2 показано, что процесс установления соответствия не сводится просто к операции над интенсивностями зрительного массива. Тем самым на объем обработки, которая должна предшествовать операции сопоставления, налагается некоторое ограничение снизу. В разд. 1.3 исследуется проблема «верхней границы». Показано, что установление соответствия между структурированными формами происходит не в целом, на основе их подобия друг другу, а

на основе сопоставления друг другу малых компонент этих форм. Таким образом, данные, изложенные в гл. 1, позволяют заключить, что в процессе того, как при предъявлении двух последовательных кадров возникает ощущение движения, каждый кадр анализируется вплоть до уровня опорных символов соответствия. Затем искомое соответствие между этими опорными символами устанавливается с помощью процесса, обсуждаемого в гл. 2. Здесь изучаются общая природа вычисления соответствия и некоторые связанные с этим вопросы.

В разд. 2.1—2.3 исследуются принципы, управляющие установлением соответствия, начиная с конфигураций, достаточно простых для того, чтобы избежать затруднений, связанных с обилием взаимодействий. Показано, что в зрительной системе человека имеется некоторая мера подобия друг другу опорных символов соответствия, называемая *сродством* (см. разд. 2.2). Соответствие между элементами определяется по исходным значениям сродства через локальные конкурирующие взаимодействия (см. разд. 2.3). В дальнейшем эти принципы объединяются и образуют *схему состязаний* для вычислений соответствия. В разд. 2.4 проверяется приложимость схемы состязаний к разнообразным конкретным случаям. Вычисленное соответствие и зависимость его от МСИ оказываются согласованными с данными о восприятии движения человеком. Среди приводимых примеров — конфигурация Тернуса, для которой имеются два способа установления зрительного соответствия: по соседнему элементу и по связанному элементу. На основе схемы состязаний предлагается объяснение такой бистабильности, включая и объяснение перехода от первого способа ко второму при возрастании МСИ.

В разд. 2.5 рассматривается вопрос о том, является ли расстояние, определяющее сродство между опорными символами соответствия, двумерным неинтерпретируемым в смысле реального физического пространства или же оно трехмерно и пространственно интерпретируемо. Хотя гл. 1 и дает основания ожидать, что это расстояние пространственно неинтерпретируемо, существуют некоторые факты, основанные на экспериментах по «гладкости движения», указывающие на обратное. Выводы, вытекающие из этого раздела, таковы: 1) процесс установления соответствия руководствуется пространственно неинтерпретируемым расстоянием; 2) гладкость движения не может служить средством измерения степени сродства.

В разд. 2.6 указывается, что наличие нижеуровневого автономного процесса установления соответствия сулит потенциальные преимущества и для других видов активности, связанных со зрением. Обсуждается возможность некоторого участия процесса установления соответствия в первоначальном формировании представлений об объекте. Выдвигается предположение, что восприятие объектов в движении не зависит от представления о неподвижном объекте, а, наоборот, может служить основой для создания такого представления.

В гл. 3 проведено исследование процесса установления соответствия с вычислительной точки зрения. Развивается теория установления соответствия при движении, основанная на понятии минимального отображения. В разд. 3.1 — 3.5 определяется конечная цель этого процесса и разрабатывается практически применимая схема, решающая задачу сопоставления на основе некоторой встроенной меры сродства и локальных взаимодействий, изученных в гл. 2. Эта схема, названная *минимальным отображением*, отыскивает наиболее правдоподобное сопоставление между наборами опорных символов соответствия. В разд. 3.6, 3.7 указанное минимальное отображение сравнивается с процессом установления соответствия зрительной системой человека. Показано, что из принципа минимального отображения можно вывести существование разных важных явлений, в том числе явлений обнаружения сдвигов, симметрии, множественной устойчивости определенных конфигураций, а также свойства минимального покрытия. В разд. 3.8 схема минимального отображения обобщается на случай произвольных опорных символов соответствия при непрерывном движении. В разд. 3.9 обсуждение процесса установления соответствия завершается рассмотрением некоторых связей между эмпирическим исследованием (см. гл. 2) и вычислительным анализом (см. гл. 3).

## ГЛАВА 1

### ПРОБЛЕМА ОПОРНЫХ ЭЛЕМЕНТОВ

#### 1.1. Теоретические соображения

В основе способности зрительной системы человека к восприятию движения лежит возможность идентифицировать различные элементы входного зрительного массива как элементы, представляющие один и тот же физический объект. Для данного элемента  $X$  в  $A(t)$  (зрительном массиве в момент  $t$ ) должно быть определено местоположение соответствующего ему элемента  $X'$  в  $A(t')$  — зрительном массиве в более поздний момент  $t'$ . Элемент  $X'$  не обязательно должен быть похож по виду на  $X$ , и на самом деле различие между ними может послужить материалом для последующего анализа движения или изменения, относимого к тому объекту, который представляют элементы  $X$  и  $X'$ .

Перед тем как изучать фундаментальную функцию установления соответствия, сопоставляющую элементы на последовательных изображениях, следует заняться более простой задачей нахождения области определения и области значений этой функции, т. е. указания множества тех элементов, которые подвергаются отображению в процессе восприятия движения. Назовем основные, опорные элементы, составляющие это множество, опор-

ными символами соответствия. При формулировке проблемы опорных элементов в терминах области определения и области значений функции эта проблема выглядит обманчиво безобидной. Тем не менее именно отсутствие единого мнения о том, каковы опорные элементы, объясняет более существенные расхождения различных теорий восприятия движения. Разнообразные подходы к зрительному анализу движения отличаются друг от друга теми предположениями, которые в них делаются относительно природы опорных элементов. Поскольку каждый конкретный их выбор обычно содержится в теории в скрытом виде, а не формулируется явно, то он редко обосновывается и следующие из этого выводы редко обсуждаются. Например, с точки зрения крайнего эмпиризма, люди могут научиться ассоциировать объекты с их проекциями (видами) и благодаря этому могут узнавать объекты в движении. Проблема принятия этой теории или отказа от нее равносильна вопросу о том, являются ли в действительности эти заученные проекции объектов опорными символами соответствия. Подобным же образом спор о том, основан анализ движения объекта на узнавании или нет, является дискуссией о том, представляют ли собой узнанные объекты (или описания объектов) опорные элементы функции соответствия.

Проблема опорных элементов должна быть рассмотрена в первую очередь вследствие ее фундаментальной важности и в конечном счете того влияния, которое она должна оказать на ход исследования. Если, к примеру, распознавание объекта действительно предшествует анализу движения, то процедура установления соответствия становится довольно простой: конкретный элемент на изображении  $A(t')$ , соответствующий, скажем, белому кролику, сидящему под дубом на изображении  $A(t)$ , определяется, вероятно, единственным образом и без особых трудностей. С другой стороны, в таких теориях восприятия становятся загадочными и установление соответствия, и трехмерная интерпретация незнакомых движущихся объектов, особенно, как это будет обсуждаться в гл. 4, в случае тех объектов, двумерные проекции которых не могут быть узнаны. Однако если опорные элементы являются более примитивными, такими как линии, точки и фрагменты краев, то известность и узнаваемость перестают играть важную роль. Тогда трудной становится проблема установления соответствия, поскольку на изображении  $A(t')$  может быть много кандидатов для возможного сопоставления, например, с отрезком прямой на  $A(t)$ .

Цель данного раздела — предложить, исходя из теоретических соображений, некоторое целесообразное множество опорных элементов. Будут приведены доводы в пользу того, что уровень, на котором осуществляется процесс установления соответствия, следует непосредственно за уровнем, на котором производится первоначальная организация исходного поля интенсивностей в содержательные единицы. (Содержательные единицы — это, как обсуждалось во введении, символы представления, смысловое содер-



жание которых определяется окружающей обстановкой\*, а не массивом интенсивностей.) В следующих двух разделах приводятся свидетельства в пользу этого взгляда. В разд. 1.2 будет показано, что организация исходных интенсивностей в содержательные единицы предшествует операции установления соответствия, а в разд. 1.3 — что эта предварительная организация ограничена образованием элементарных единиц.

**Разумное определение множества опорных элементов.** Обычный подход к задачам сопоставления изображений, содержащийся как в психологической литературе [Кабриски, 1966; Анстис, 1970; Белл и Лаппен, 1973; Пэнтл и Пиччиано, 1976], так и в литературе, посвященной приложениям [Лиз, Новак и Тэйлор, 1970; Смит и Филлиэс, 1972; Волфертс, 1974], заключается в принятии предположения о том, что процесс установления соответствия происходит ранее любой организации исходных интенсивностей\*\*.

Этот общий взгляд можно далее подразделить, основываясь на конкретных операциях, используемых для установления соответствия. При одном из подходов [Анстис, 1970] отдельные точки связываются между собой на основе сходства их интенсивностей. В этом случае роль опорных символов соответствия берут на себя одиночные точки массива интенсивности. При более общих подходах в качестве опорного элемента рассматривается некоторая область  $A$  данного массива интенсивности, а соответствующая ей область  $A'$  на втором массиве интенсивности ищется исходя из сходства распределений интенсивности в них. Степень сходства этих массивов обычно измеряется либо с помощью метода взаимной корреляции, при котором ищется максимальная корреляция, либо с помощью вычитания<sup>1.1</sup>, когда нужное сопоставление определяется по минимальному значению разности.

В указанных подходах соответствие устанавливалось на основе сходства между распределениями интенсивности. Против установления соответствия с помощью таких полутоновых сравнений сходства имеются два основных возражения.

Во-первых, можно ожидать, что полутоновые корреляции дадут точное соответствие только в очень простом случае параллельного переноса в плоскости изображения. В общем случае, когда подвергаемые сравнению картинки представляют объект, движущийся произвольным образом, нет никаких оснований полагать, что какое-либо из указанных полутоновых сравнений должно давать надежный результат<sup>1.2</sup>. Одна из возникающих при этом проблем — проблема «размера окна». Сравнение интенсивности не может быть выполнено ни над одиночными точками, ни над изображением в целом. Корреляция должна быть установлена меж-

\* То есть рассматриваемым объектом — *Прим. ред.*

\*\* В предисловии к русскому переводу сборника «Психология машинного зрения» (М.: Мир, 1978) показано, что при таком подходе возникают непреодолимые трудности, и отмечается, что методы, характерные для искусственного интеллекта, подсказывают, что сопоставление различных проекций объекта должно производиться на уровне их символьных описаний. — *Прим. ред.*

ду частями изображения нужного размера, но не существует никакого удовлетворительного способа для предварительного выбора этого размера. В случае нескольких объектов, одновременно участвующих в различных движениях, ситуация становится более сложной и в еще меньшей степени поддается решению полутонными сравнениями.

Во-вторых, полутонные распределения и их изменения не связаны непосредственно с физическими объектами и их движением, тогда как именно движение и должно быть определено. Например, как это иллюстрируется иллюзией Корнsvита [Корнsvит, 1970; Ратлифф, 1972], наблюдаемый «край» может характеризоваться совершенно различными распределениями интенсивности, которые с точки зрения восприятия не отличимы друг от друга. Иллюзия Корнsvита не должна удивлять; при различных условиях освещения и ориентации данный физический край объекта может порождать различные распределения интенсивности, сами по себе не представляющие никакого интереса для зрительной системы человека или воспринимающего устройства, которые должны восстановить физическую структуру окружающих объектов (но же обсуждение, касающееся вычисления стереоскопической диспаратности, см. в работах [Марр, 1974; Марр и Поджио, 1976]). Когда край обнаружен, то образующее его распределение интенсивности может быть заменено символическим представлением края и после этого соответствие может устанавливаться между двумя представлениями края. Проведенное в разд. 1.2 рассуждение послужит для построения контрпримера к рассматриваемой гипотезе полутонной корреляции.

Проведенное обсуждение наводит на мысль, что анализ движения должен выполняться только после того, как поле исходных интенсивностей организовано в некоторую совокупность единиц. Единицы, по всей вероятности, выделяются и организуются иерархическим образом, т. е. сначала отыскиваются такие единицы, как фрагменты краев, полосы и малые пятна, которые затем организуются в более структурированные формы и в конечном итоге в конкретные объекты.

Если считать такую схему организации справедливой, то подходящими кандидатами на роль опорных символов соответствия являются единицы, расположенные вблизи нижнего уровня иерархии, и поэтому мы будем их называть единицами *нижнего уровня*. Обоснованность этой схемы зависит частично от того способа, которым осуществляется процесс установления соответствия, и может быть правильно понята только с учетом содержания последующих разделов. Суть этого обоснования, однако, такова. Проблема, с которой сталкивается зрительная система при установлении соответствия, заключается в определении вероятности того, что элементы  $X$  и  $X'$  являются одним и тем же объектом, находящимся в состоянии движения. Следовательно, нужна какая-то мера правдоподобия того факта, что  $X$  и  $X'$  отображают один и тот же объект, но после небольшого его перемещения. Не-

возможно себе представить, чтобы в памяти хранились разнообразные фигуры вместе с соответствующими им мерами правдоподобия; значит, эта мера должна вычисляться. И лишь для вполне определенного класса единиц, а именно для членов совокупности опорных элементов действительно имеется «хранящаяся» мера правдоподобия, которую будем называть *сродством*\*. Для более сложных рисунков соответствие вычисляется по значениям сродства их составных частей с учетом возникающих взаимодействий типа тех, что описаны в разд. 2.1. Поэтому можно ожидать, что опорные элементы являются теми строительными блоками, из которых можно создавать сложные конфигурации. В следующих двух разделах обосновывается эта точка зрения. В них содержатся данные, свидетельствующие о том, что соответствие действительно устанавливается сопоставлением опорных единиц, таких как края, отрезки линий и малые пятна. В этом контексте целесообразно ввести понятие *первичного эскиза*, определенного Марром в его теории предварительной зрительной обработки [Марр, 1976]. Первичный эскиз представляет собой множество опорных единиц, которые первыми формируются в процессе зрительного анализа и в дальнейшем служат строительными блоками для конструкций более высоких порядков. На основе проведенного обсуждения можно ожидать, что множество опорных символов соответствия будет практически эквивалентно совокупности элементов, составляющих первичный эскиз. Кажется, это действительно так: наши два исследования в поисках опорных единиц, по-видимому, приводят к схожим множествам элементов.

## **1.2. Установление соответствия не является полутоновой операцией**

В предыдущем разделе утверждалось, что полутоновые операции неадекватны задаче определения движения и что организация исходных данных в элементарные содержательные единицы должна предшествовать процессу установления соответствия.

В данном разделе приводится демонстрационный эксперимент, свидетельствующий в пользу этого утверждения. Эксперимент основан на видимом движении от картинке к картинке. Эти картинки задуманы таким образом, что полутоновые операции подразумевают один вид движения, в то время как схема, основанная на предварительной организации данных в содержательные единицы, предсказывает движение другого вида.

Профили интенсивности *A* и *B* двух картинок, использованных в эксперименте, изображены на рис. 1.1. Оба профиля полу-

---

\* В оригинале *affinity* — близость, сходство, взаимная склонность. Мы переводим это слово термином «сродство». — *Прим. пер.*

ченны из кривой  $S$ . Профиль  $A$  получается из  $S$  путем «сглаживания» правой ступеньки, а профиль  $B$  — левой. С позиции восприятия  $S$  содержит два резких края в точках  $p$  и  $q$ , тогда как  $A$  имеет единственный резкий край в точке  $p$ , а  $B$  — в точке  $q$ . Для дальнейшего объяснения нужно ввести определение положения  $B$  относительно  $A$ . Пусть координате, при которой  $A$  и  $B$  налагаются друг на друга, соответствует 0. Положительная координата будет означать, что  $B$  было смещено вправо, отрицательная — вле-

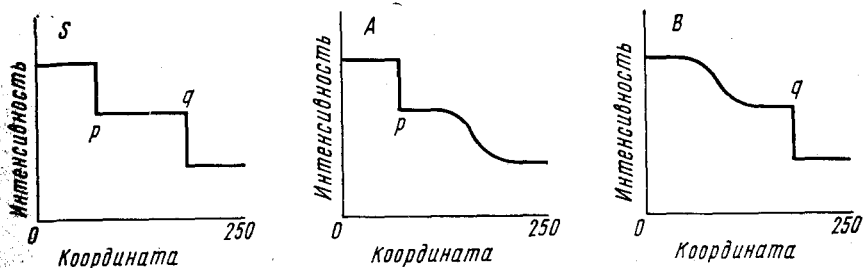


Рис. 1.1. Профили интенсивности:  $S$  — с двумя резкими краями;  $A$  — с единственным резким краем в точке  $p$ ;  $B$  — с единственным резким краем в точке  $q$  по. Вся картинка, как показано на рис. 1.1, имеет размер по горизонтали 250 единиц. Подразумевается, что единицы измерения координаты и интенсивности берутся только в относительном масштабе; их абсолютные значения могут изменяться в широких пределах.

Какое движение должно возникать при попеременном предъявлении этих двух картинок? Если соответствие устанавливается между воспринимаемыми краями, то предсказание несложно: края  $p$  и  $q$  должны быть видны в движении. Если сопоставление основывается на полутоновой корреляции, то предсказание будет иное. На рис. 1.2 представлена функция взаимной корреляции

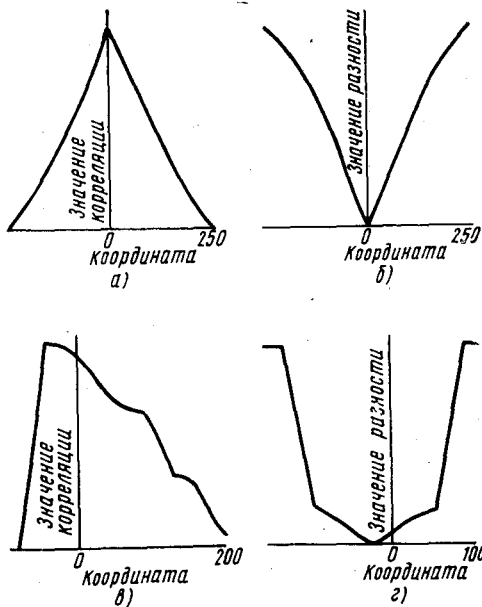


Рис. 1.2. Результаты определения степени подобия между профилями  $A$  и  $B$  на рис. 1.1 (отложена по вертикальной оси) как функции их взаимного расположения (отложено по горизонтальной оси): использованы глобальная взаимная корреляция (а), процесс глобального вычитания (б), локальная взаимная корреляция (в), процесс локального вычитания (г)

между профилями интенсивности  $A$  и  $B$ . Как видно, взаимная корреляция достигает своего максимума в точке 0. Значит, предсказывается следующее: 1) если  $A$  и  $B$  показываются точно совмещенными друг с другом (координата 0), то не должно возникать никакого движения; 2) если первым показывается  $A$ , а затем  $B$ , сдвинутое, скажем, на  $-20$  единиц, то ожидается движение на 20 единиц влево, так как при таком перемещении будет максимизироваться взаимная корреляция.

Таким образом, можно провести критический эксперимент, в котором два описанных метода предсказывают противоположные результаты. Когда две указанные картинки предъявлялись тахистоскопически, то наблюдаемое движение происходило между видимыми краями, что противоречит предсказанию, основанному на взаимной корреляции. Время предъявления составляло от 100 до 200 мс с межстимульным интервалом от 30 до 70 мс. Угловой размер каждой картинки составлял  $3,5^\circ$  зрительного угла, а угловое расстояние между двумя упомянутыми краями —  $1,5^\circ$ .

Для сравнения картинок предлагались помимо взаимной корреляции другие полутоновые операции. На рис. 1.26 показаны результаты применения к обсуждаемым профилям второго метода, называемого «операцией вычитания»<sup>1,3</sup>. В этой операции сопоставление определяется по минимальному значению кривой на рис. 1.26. Как и взаимная корреляция, сопоставление оптимально в точке 0. Следовательно, предсказания метода, основанного на вычитании, эквивалентны предсказаниям, основанным на методе взаимной корреляции, и точно так же опровергаются результатами эксперимента. Другие полутоновые операции, такие как локальная корреляция (рис. 1.2в) и локальное вычитание (рис. 1.2г), были изучены и отвергнуты подобным образом по той же основной причине: изменения в распределениях исходной интенсивности не отражают непосредственно изменений наблюдаемой ситуации. Поэтому предварительным условием определения физического движения на основе изменяющегося оптического массива является организация входного зрительного массива в единицы, соответствующие определенным физическим элементам.

На вывод о том, что установление соответствия при движении основано на сопоставлении опорных символов, а не интенсивностей, в случае малых перемещений должно налагаться следующее ограничение. Имеются указания на существование процесса обнаружения движения, откликающегося на изменения в распределениях интенсивности [например, «обращенное фи-движение», открытое Анстисом (1970), и процесс с малой зоной действия у Брэдика (1974)]. Этот процесс характеризуется малой зоной действия (15—20 угл. мин) и тем, что он более эффективен в периферийном зрении, в противоположность установлению соответствия между символами, которое имеет большую зону действия и эффективно прежде всего в центре поля зрения. Если действительно существуют два таких процесса, то вполне возможно, что они могут служить для разных целей. Периферийный процесс,

основанный на интенсивности, подходит для «системы раннего повешения», обнаруживающей изменения и направляющей внимание. Он также может быть полезен при нахождении границ разрывов, где происходит скачкообразное изменение скорости движения элементов зрительного поля. В то же время, процесс установления соответствия в результате нахождения и прослеживания движения опорных символов является инструментом поддержания перцептивной целостности движущихся объектов. Кроме того, он необходим на этапе трехмерной интерпретации, что подробно будет рассмотрено в гл. 4. Таким образом, для целей раннего исследования возможный существующий процесс обнаружения, основанный на анализе интенсивностей, имеет второстепенное значение и далее обсуждаться не будет.

В заключение скажем, что результаты, полученные в этом разделе, указывают нижнюю границу объема обработки, предвещающей установление соответствия между элементами изображений. В следующем разделе будет рассмотрена проблема верхней границы. Исходя из теоретических предпосылок нами будет сделан вывод, что в иерархии единиц, организуемых зрительной системой, опорные элементы соответствия находятся, вероятно, на том нижнем уровне или близ него. В следующем разделе будут представлены экспериментальные свидетельства в пользу такого взгляда.

### **1.3. Опорные символы соответствия не являются структурированными формами**

В этом разделе будут описаны пять демонстрационных экспериментов в поддержку того взгляда, что процесс установления соответствия не основывается на развитом восприятии форм. Будут приведены аргументы в пользу того, что воспринимаемое соответствие между движущимися структурированными формами устанавливается не между целыми формами на основе их подобия друг другу, а является результатом сопоставления простых составляющих этих структурированных форм. Каждый эксперимент будет описан с последующим кратким обсуждением результатов.

**Демонстрационный эксперимент 1: «разбитое колесо».** Демонстрация с использованием изображения «разбитого колеса» напоминает хорошо известный кинематографический эффект, называемый иногда явлением «колеса повозки», при котором колесо со спицами выглядит вращающимся в направлении, противоположном истинному направлению вращения. Это явление указывает на склонность зрительной системы к выбору того из двух возможных сопоставлений, которое вызывает минимальное изменение (в данном случае угловое). В рамках рассмотрения проблемы опорных элементов явление «колеса повозки» допускает две различные интерпретации:

1. Сперва происходит организация малых единиц в законченную форму (колесо), а затем форма на первом изображении  $A(t)$  сопоставляется с формой,

найденной на последующем изображении  $A(t')$ . Поскольку существует несколько способов их полного взаимного сопоставления, то выбирается тот, который обеспечивает минимальное изменение.

2. Соответствие устанавливается между малыми единицами, образующими колесо, ощущение движения всей формы возникает на более поздней стадии, вытекая из движений составных частей.

В случае явления «колеса повозки» эти два разных метода анализа дают одинаковые результаты. Изображение «разбитое колесо» было сконструировано таким образом, чтобы две гипотезы имели различные следствия. В данном эксперименте вращающаяся фигура изображает колесо, в котором каждая вторая спица сломана, так что средняя часть ее отсутствует (рис. 1.3а). Пусть  $\alpha$  — угол между двумя соседними спицами, и будем считать, что от одного изобра-

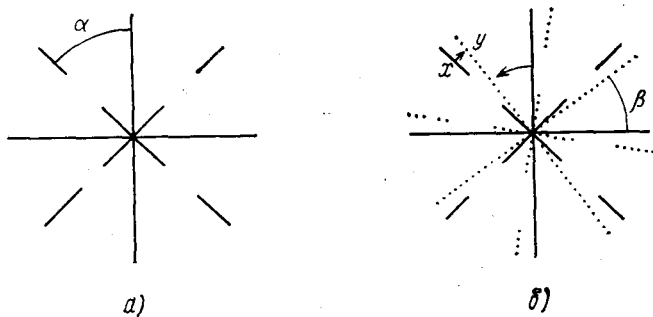


Рис. 1.3. Демонстрационный эксперимент «разбитое колесо»: — — первый кадр; - - - - второй кадр

жения до другого колесо поворачивается на  $\beta^\circ$  против часовой стрелки. Рассмотрим теперь, что произойдет при  $\alpha/2 < \beta < \alpha$  (на рис. 1.3б  $\alpha$  составляет  $12^\circ$ , а  $\beta$  —  $8^\circ$ ). Если рассматривать фигуру в целом, то наилучшее сопоставление достигается поворотом колеса на первом изображении на  $\beta^\circ$  против часовой стрелки. Однако если бы короткий отрезок ( $x$  на рис. 1.3б) рассматривался как опорный элемент и искался бы его ближайший эквивалент на другом изображении, то можно было бы выбрать отрезок по направлению часовой стрелки ( $y$  на рис. 1.3б). С первой точки зрения такой выбор невозможен, но со второй — очень приемлем (хотя по причинам, которые будут обсуждаться в подразд. 2.4.2, это не обязательно так). Результат эксперимента таков: при соответствующим образом подобранных временах (для получения гладкого движения в условиях темной комнаты требовалось время предъявления 50 мс и МСИ 30 мс) колесо распадается на три четко видимых кольца. Из них внутреннее и внешнее вращаются по часовой стрелке, а среднее кольцо — против часовой стрелки. Это распадение указывает на то, что движение возникает не между полными формами, а между их составными частями.

**Демонстрационный эксперимент 2: «поезд из кубиков».** «Поезд из кубиков» составлен, как показано на рис. 1.4а, из «вагонов» с окнами. Соседние вертикальные линии удалены друг от друга на  $x$  единиц, а сам поезд от изображения к изображению сдвигается на  $y$  единиц вправо (рис. 1.4б). При  $x/2 < y < x$

В эксперименте использовались значения  $x=0,4^\circ$  и  $y=0,3^\circ$  зрительного угла (при времени предъявления 50 мс и МСИ 30 мс) есть два основных вида восприятия «движущегося поезда». В первом случае фигура может расщепиться: «окна» движутся влево, тогда как остальная часть поезда движется вправо. Во втором случае весь поезд может двигаться вправо. Первый из этих видов восприятия похож на явление разбитого колеса с заменой вращения на линейный перенос. Этот вид восприятия реализуется всякий раз, когда наблюдатель

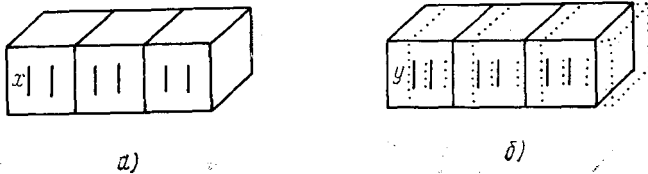


Рис. 1.4. Демонстрационный эксперимент «поезд из кубиков»

проследит взгляд на неподвижной точке экрана и не позволяет глазам проследить движущийся поезд. Второй вид восприятия является, вероятно, результатом прослеживающих движений глаз. Предположим, что наблюдатель проследит фигуру абсолютно точно, т. е. «поезд» не движется относительно его глаза. Тогда из того факта, что картинка осталась неизменной, но для поддержания стабильности изображения потребовались прослеживающие движения глаз, следует, что вся фигура в целом подверглась переносу.

**Демонстрационный эксперимент 3: «неоднородная звезда».** Используется восьмиугольная звезда, у которой лучи через один сделаны шире (второй вариант — длиннее). Пусть опять  $\alpha$  будет углом между соседними лучами, а  $\beta$  — углом поворота (рис. 1.5a). Если  $\beta = \alpha/2$ , то «неоднородная звезда» вращается в том направлении, в каком каждому широкому лучу сопо-

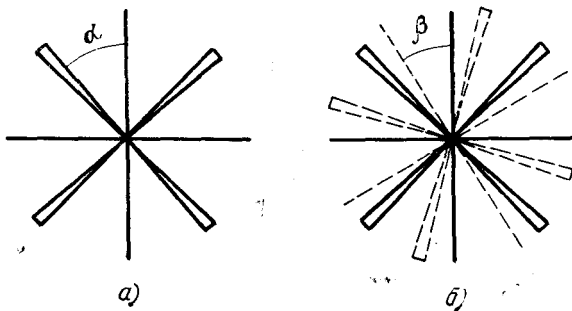


Рис. 1.5. Демонстрационный эксперимент «неоднородная звезда»

ставляется соседний широкий луч, а каждому узкому — соседний узкий. Если  $\alpha < \beta < 2\alpha$ , то звезда воспринимается вращающейся в «неправильном» направлении, при этом каждый узкий луч становится шире, а каждый широкий — уже.



В изображении «неоднородная звезда» фигура не расщепляется на меньшие единицы, поэтому значимость этого эксперимента для обсуждаемой проблемы можно подвергнуть сомнению на том основании, что эксперимент вполне может быть объяснен, если рассматривать эту фигуру как целое. Естественно рассуждать так: фигура в целом имеет два возможных варианта сопоставления — тождественное сопоставление при повороте на  $\beta^\circ$  назад и другое сопоставление, пространственно более близкое ( $\alpha^\circ - \beta^\circ$ ), но вызывающее изменения фигуры. Таким образом, можно предложить построить метрическое пространство, основывающееся как на пространственном расстоянии, так и на подобии, в котором результат второго сопоставления будет «ближе» к оригиналу по сравнению с первым.

Это возражение неубедительно по следующей причине. Главное преимущество проведения анализа формы до проведения операции сопоставления заключается в возможности идентифицировать две фигуры как соответствующие друг другу на основе похожести фигур. Поэтому можно было бы ожидать, что две сложные, идентичные и близко расположенные друг к другу фигуры неизбежно должны были быть приведены в соответствие, что, однако, противоречит описанным экспериментальным результатам. Далее, если полный анализ формы действительно предшествует процессу установления соответствия, то можно было бы ожидать, что хорошее соответствие сложных фигур должно быть более сильным указателем их сопоставимости, чем соответствие друг другу небольших и простых составных частей этих фигур. Однако эксперименты с одиночными лучами показывают, что для изменения выбираемого направления движения на противоположное как одиночному лучу, так и всей звезде требуется одно и то же значение отношения  $\beta/\alpha$ .

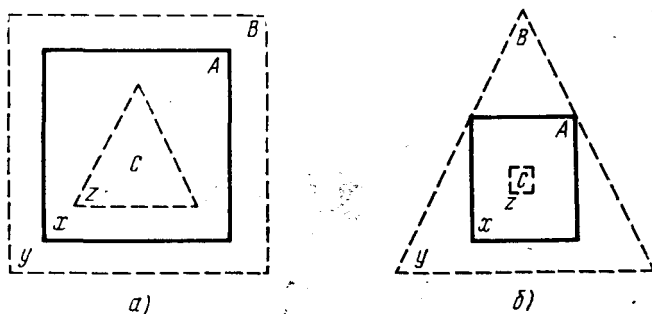
**Демонстрационный эксперимент 4: «вращающаяся спираль».** В хорошо известной иллюзии вращающаяся спираль (при непрерывном или дискретном предъявлении) выглядит расширяющейся или сжимающейся в зависимости от направления вращения [Колерс, 1966]. Если концевые точки спирали скрыты от наблюдателя, то воспринимаются только ее движения наружу и внутрь, а вращение не воспринимается [Уоллок, Вайс и Адамс, 1956]. При этом спираль в целом участвует только во вращательном движении. Однако если рассматривать в качестве опорных элементов малые участки спирали, то само собой напрашивается одно правдоподобное объяснение этой иллюзии. Действительно, установление соответствия между малыми частями спирали и их ближайшими соседями подразумевает наличие движения, перпендикулярного вращению. Если рассматривать эту картину через узкую радиальную щель, то можно наблюдать движение наружу и внутрь, порожденное установлением локального соответствия.

**Демонстрационный эксперимент 5: «установление соответствия и форма».** В прошлом были предприняты различные попытки выяснить то влияние, которое сходство формы оказывает на воспринимаемое соответствие фигур. Следуя предположению о том, что процесс сопоставления должен отдавать предпочтение сопоставлению похожих фигур, Колерс (1962) сравнивал гладкость воспринимаемого движения для фигур схожих и несхожих. Поскольку степени гладкости оказались одинаковыми и не зависели от сходства фигур, то Колерс делал вывод о том, что для зрительной системы все двумерные фигуры в равной степени похожи друг на друга. С другой стороны, имелись некоторые данные [Орлански, 1940; Фрисби, 1972] о том, что для простых стимулов, в особенно-

и отрезков прямых различной ориентации, имело место некоторое влияние сходства на «оптимальность» воспринимаемого движения.

Данные о том, что подобие сложных форм не влияет каким-либо явным образом на установление соответствия между ними, по-видимому, согласуются с выдвинутой в этом разделе идеей о том, что процесс сопоставления предшествует организации опорных единиц в структурированные формы. Эти данные, однако, нельзя считать относящимися к обсуждаемой проблеме. Главная причина заключается в том, что данные [за исключением результатов Навона (1976) и некоторых экспериментов Колерса (1972)] были основаны на оценках скорости движения, которые, как это показано в разд. 2.4, не являются надежной мерой тенденции фигур «к слиянию», для измерения которой эти оценки были предназначены.

Прямым методом определения влияния подобия фигур на процесс сопоставления является метод, который назовем методом *конкурирующих движений*. При использовании этого метода попеременно предъявляются два кадра. Первый кадр содержит один элемент (или фигуру), а второй — два элемента. Возникает вопрос: «Участвует фигура на первом кадре в движении с первым элементом второго кадра или со вторым?». На рис. 1.6а приведен



1.6. Установление соответствия между похожими (а) и непохожими (б) формами

пример экспериментального изображения, демонстрирующего конкурирующие движения. На первом кадре представлен один квадрат  $A$ , а на втором — внешний квадрат  $B$  и внутренний треугольник  $C$  (время предъявления 120 мс, МСИ 40 мс). При предъявлении этих изображений воспринимается соответствие  $A \leftrightarrow B$ , т. е. предпочтительным является движение между двумя квадратами. В противоположность некоторым выводам, сделанным другими авторами, полученные результаты указывают на то, что фигуры действительно обладают различной тенденцией к слиянию. Но свидетельствует ли это предпочтение о влиянии, оказываемом подобием фигур? Не обязательно. Если проверить подругим отдельные линии, составляющие изображение, то предпочтение остается тем же самым. К примеру, если  $x$  на рис. 1.6а показывается в условиях конкурирующего движения с  $y$  и  $z$ , то

предпочитается его движение в сторону  $y^*$ . Стремление квадрата  $A$  слиться с квадратом  $B$ , а не с треугольником  $C$  можно, таким образом, объяснить на основе движения составляющих их элементов. При этом нет нужды предполагать, что опорными элементами являются формы целиком или что сходство форм определяет установление соответствия, наблюдаемого в эксперименте. Если в данном примере установленное соответствие не противоречит ни сходству фигур, ни локальному сопоставлению элементов, то в следующем примере из подобия и локального сопоставления следуют разные выводы.

На рис. 1.6б наблюдается установление соответствия между *непохожими* фигурами: предпочитается сопоставление прямоугольника  $A$  треугольнику  $B$ , а не внутреннему прямоугольнику  $C$ . Это предпочтение, как и раньше, согласуется с установлением соответствия по составляющим элементам, например сопоставление  $x \leftrightarrow y$  более предпочтительно, чем  $x \leftrightarrow z$  \*\*. Таким образом, процессом сопоставления руководит движение составных элементов, а не подобие полных форм. Подводя итоги обсуждению эксперимента 5, можно сказать, что: 1) в противоположность прежним представлениям различные фигуры действительно отличаются друг от друга своей тенденцией к слиянию, причем предпочтительное сопоставление согласуется с тем движением, которое устанавливается между их составными частями; 2) нет никаких указаний на то, что структурированные фигуры в целом являются частью множества опорных элементов или что процесс установления соответствия основывается на сходстве фигур.

Все описанные выше демонстрационные эксперименты поддерживают положение о том, что операции установления соответствия вовсе не должен предшествовать сколь-нибудь подробный анализ формы и что движение сложных фигур складывается из движения их составных частей.

В двух из нижеследующих разделов книги приводятся дополнительные обоснования этого взгляда и его разработка. Во-первых, обсуждение некоторых связей между процессом сопоставления и структурированными фигурами отложено до подразд. 2.4.2, так как они рассматриваются в свете схемы установления соответствия, вводимой в гл. 2. Во-вторых, в разд. 2.5 показывается, что операция установления соответствия не зависит от трехмерной интерпретации сцены, и, таким образом, получают поддержку тот взгляд, что по своей природе установление соответствия является операцией нижнего уровня<sup>1.4</sup>.

---

\* В данном случае  $x$ ,  $y$ ,  $z$  представляют собой малые окрестности соответствующих вершин квадратов и треугольников. — *Прим. ред.*

\*\* В рамках отдельного эксперимента с локальными компонентами, который, как и в предыдущем примере, автором детально не описывается. — *Прим. ред.*

## ПРОЦЕСС УСТАНОВЛЕНИЯ СООТВЕТСТВИЯ

## 2.1. Общая схема

## 2.1.1. Введение

В гл. 1 был сделан вывод о том, что в процессе движения соответствие устанавливается в результате сопоставления элементарных опорных символов соответствия. В гл. 2 и 3 мы обращаемся к вопросу о том, как устанавливается это соответствие. В данной главе рассматриваются общие принципы, лежащие в основе процесса сопоставления. Изучение некоторых простых конфигураций в движении показывает, что в зрительной системе применяется определенная мера сродства опорных символов, которую можно приближенно считать мерой их подобия друг другу. Соответствие между наборами опорных символов устанавливается по делениям степени сродства этих символов с помощью локально циркулирующих взаимодействий.

Общий вывод, который можно сделать из гл. 2, состоит в том, что установление соответствия является автономным процессом нижнего уровня. Оно зависит от пространственно-временных параметров опорных символов изображения и параметров их сходства, а не от процессов более высокого уровня, таких как распознавание и трехмерная интерпретация. Такой автономный низкоразрядный процесс установления соответствия, по-видимому, дает возможность объяснить то, как различные изображения движущегося или меняющегося объекта идентифицируются зрительной системой как представления одного и того же объекта. Может возникнуть впечатление, что при наблюдении за движущимся или меняющимся объектом зрительная система иногда может выйти за пределы имеющейся в ее распоряжении входной информации. Вид движущегося или изменяющегося объекта (при детальном предъявлении) может очень сильно меняться от кадра к кадру. Однако даже тогда, когда объект нам незнаком, зрительная система часто в состоянии идентифицировать меняющиеся картинку как изображения одного и того же объекта, находящегося в состоянии движения или изменения. Поскольку единственной информацией, доступной зрительной системе, является внешний вид объекта, то способность системы к такой идентификации на основе принципиально различных изображений выглядит загадочной. Эта способность действительно удивительна, если сопоставление производится на основе прошлого опыта, распознавание изображенных объектов или на основе подобия друг другу наблюдаемых изображений в целом. Но если производится сопоставление элементарных опорных символов с помощью описанной в данной главе схемы, то предыдущий опыт, распознавание или сходство изображений в целом уже не являются необходимыми

условиями восприятия различных изображений как представлений одного и того же объекта.

Возможность того, что соответствие между различными изображениями строится из попарного соответствия между собой элементарных единиц, в прошлом неоднократно подвергалась критике. Поскольку, на мой взгляд, этот подход надо возродить, то вкратце рассмотрим и подвергнем критическому анализу основные аргументы, опровергающие его<sup>2.1</sup>.

Проблема установления соответствия под такими названиями, как «феноменальная идентичность» и «селективное слияние», привлекла значительное внимание психологов-гештальтистов. Фактически знаменитая работа Вертгеймера (1912), которая считается началом гештальтпсихологии, была сосредоточена на явлении кажущегося движения, включая и интересующую нас проблему установления соответствия. Работа самого Вертгеймера была продолжена исследованиями Тернуса, Метцгера, Мишотта, Метелли и других [см. Коффка, 1935; Вернон, 1966; Эллис, 1967; Арнхейм, 1974]. Эти исследователи рассматривали проблему в плане, несколько более широко, чем очерчено выше, и поэтому возникают сомнения, можно ли всю совокупность изучаемых ими явлений объединить в одну тему<sup>2.2</sup>. При обсуждении проблемы установления соответствия мы не будем рассматривать те случаи, в которых идентичность объектов устанавливается на основе сознательных рассуждений или долговременной памяти.

Одним из важнейших вкладов в исследование этой проблемы с позиций теории гештальтпсихологии была демонстрация того факта, что соответствие устанавливается между целыми полями<sup>2.3</sup> и не может быть адекватно проанализировано в терминах составных частей этих полей. Аргументация базировалась на построении примеров, где один и тот же элемент в различных полях связывался в различные пары. На рис. 2.1 приведены некоторые примеры. На всех рисунках пустые кружки представляют первое из двух предъявлений, а крестики — второе. (В реальных предъявлениях все показываемые элементы были одинаковы.) Кружок с крестиком внутри означает, что данный элемент находился на этом месте в обоих предъявлениях.

На рис. 2.1а наблюдалось установление соответствий  $A \rightarrow B$  и  $B \rightarrow C$ . На рис. 2.1б интерес представляет скорее соответствие  $A \rightarrow B$ , а не  $A \rightarrow A$ <sup>2.4</sup>. Для демонстрации того факта, что процесс сопоставления более существенно зависит от «роли», играемой элементами в поле, чем от их пространственной близости, были использованы рис. 2.1а, б. На рис. 2.1в и д устанавливается соответствие  $A \rightarrow D$ ,  $B \rightarrow E$  и т. д., тогда как на рис. 2.1г и е, аналогичных рис. 2.1в и д,  $D$ ,  $E$  и  $F$  соответствуют сами себе, т. е. они не выглядят движущимися. Вывод гештальтистов сводился к тому, что функция установления соответствия есть функция, отображающая поле на поле, и сужать ее далее нельзя. В этом доводе есть, однако, ошибка. Верно, что функция установления соответствия является *глобальной* в том смысле, что изменения в

ной части поля могут оказать влияние на то, с какими элементами связывается элемент в другой части поля. Однако из этого следует, что эта функция не может быть подвергнута дальнейшей декомпозиции и выражена через простые локальные операции. Этот довод схематически иллюстрируется рис. 2.2. На рис.

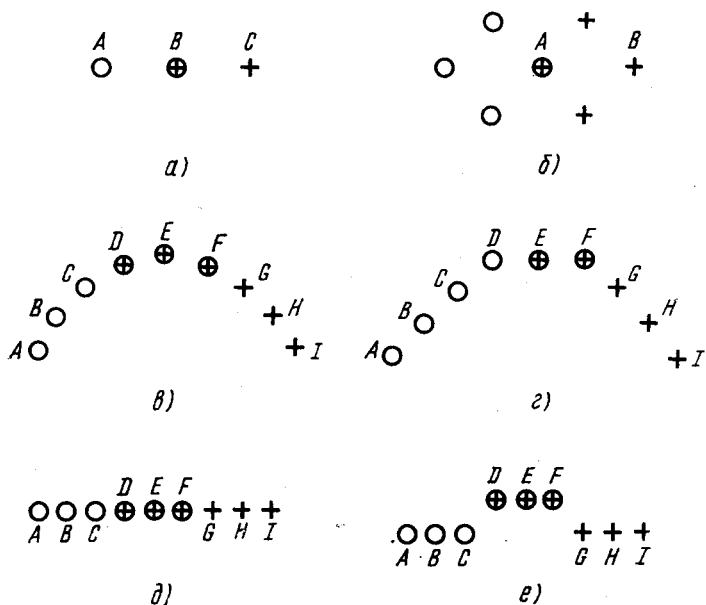
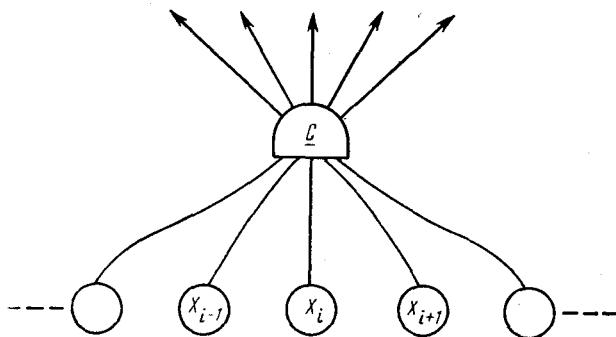


рис. 2.1. Примеры конфигураций с различным сопоставлением элементов, использовавшиеся в гештальтпсихологии

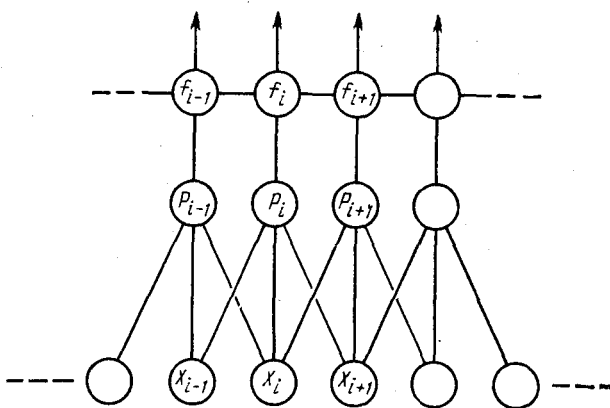
а) и б) функция установления соответствия  $C$  является  $n$ -значной вектор-функцией. Все, что можно сказать о «гештальтроне» на рис. 2.2а, — это то, что  $C = C(x_1, \dots, x_n)$ . На рис. 2.2б имеем  $C = (f_1, \dots, f_n)$ , где  $f_i = f_i(P_i, f_{i-1}, f_{i+1})$ . Обе функции глобальны в определенном ранее смысле. В первом случае, быть может, нет более простого способа изучить функцию  $C$ , чем исследовать ее введении на различных входных наборах. Во втором случае любой «выход» и является функцией всех «входов», но имеется простой метод изучения и разложения функции  $C$  с использованием локальных функций  $P_i$  и взаимодействий  $f_i$ . Из следующих разделов станет видно, что такой тип структуры, по-видимому, действительно лежит в основе реального процесса установления соответствия. Эти функции изображены на рис. 2.2 в виде сетей с учетом того, что в конечном счете они реализуются некоторой логической структурой, причем структура с локальными связями на рис. 2.2б является более правдоподобной. К этому вопросу мы вернемся в гл. 3.

Один из современных исследователей — Колерс (1972) — заново изучил проблему установления соответствия. Полученный им

результат состоял в том, что в очень широких и трудно поддающихся точному определению пределах любой стимул может связаться в пару с любым другим стимулом. К примеру, изменяя одну букву, можно ее гладко трансформировать в другую, квадрат — в круг или треугольник и т. д. Затем Колерс пытается об-



а)



б)

Рис. 2.2. Два способа организации вычисления глобальной функции: «гештальтроном» (а) и с помощью локального процесса (б)

наружить различия в той «готовности» или «легкости», с какой одна фигура преобразуется в другую. Это разумный подход: поскольку данный элемент  $X$  на изображении  $A(t)$  может быть связан почти с любым элементом  $X'$  на  $A(t')$  и поскольку в конечном итоге зрительная система выбирает некоторый конкретный элемент  $X'$ , то имеет смысл упорядочить пары элементов таким способом, который дал бы возможность предсказывать это реально устанавливаемое соответствие. Для достижения этой цели применялись главным образом два критерия: временной промежуток

гладкость движения. Критерий, основанный на временном промежутке, состоял в следующем. Как указывалось во введении, восприятие кажущегося движения зависит от временного промежутка между предъявлением стимулов. Предполагалось, видимо, что чем сильнее у данной пары тенденция к слиянию, тем в большей степени оно должно быть нечувствительно к отклонениям от оптимального временного промежутка<sup>2,5</sup>. Однако оказалось, что эта чувствительность не меняется в значительной мере при переходе от одной пары элементов к другой. Второй применявшийся критерий состоял в субъективной оценке гладкости движения, усредненной по большому числу испытуемых, однако и здесь не обнаружилось значительных различий. Колерс, как и гештальтисты, казался, но по другим причинам, от систематического анализа процесса сопоставления в терминах опорных элементов.

Что касается подхода, использующего гладкость движения как показатель тенденции к слиянию, то, применяя его, мы исходим главным образом из методологических соображений. Временные характеристики могут отражать некоторые параметры системы, имеющие почти никакого отношения к обсуждаемой проблеме. Не более важно то, что, как мы увидим в разд. 2.4, исходное предположение о гладкости не имеет под собой оснований, ибо гладкость движения *не может* служить мерой тенденции к слиянию.

Подводя итоги, можно сказать, что эксперименты гештальтистов, указывающие на то, что функция установления соответствия является глобальной, не устраняют возможности количественного анализа этой функции с помощью более простых локальных операций. Попытки начать такой анализ, в частности сравнения между собой так называемой тенденции к слиянию у пар элементов, оказались неудачными, но эту неудачу можно объяснить как результат использования неадекватных методов. Следовательно, нет никаких серьезных оснований для дискредитации выдвинутой в гл. 1 идеи о том, что установление соответствия при движении вытекает из сопоставления элементарных опорных символов.

### 2.1.2. Сродство и взаимодействия

Проблема, которую нам теперь предстоит рассмотреть, — это выявление процессом установления соответствия опорных символов на одном кадре с опорными символами на следующем кадре. Исследование этой проблемы будет состоять из двух этапов. Если установление соответствия двух сложных фигур складывается из сопоставления их элементарных составных частей, то разумно первым делом исследовать установление соответствия одиночных опорных символов или пар их, взятых отдельно. На втором этапе будут исследованы более сложные конфигурации.

Главный результат первого этапа (см. разд. 2.2) состоит в том, что установление соответствия между парами изолированных символов руководствуется определенной «встроенной» мет-



рикой взаимного сходства, называемой *сродством*. На втором этапе (см. разд. 2.3) показано, что установление соответствия между сложными фигурами определяется значениями сродства их составных частей при использовании простых локальных взаимодействий.

## 2.2. Сродство

Экспериментальные изображения, содержащие небольшое число изолированных опорных символов, иллюстрируют результат, имеющий фундаментальное значение для установления соответствия при движении: разные опорные символы с разной вероятностью связываются в пары в зависимости от разнообразных параметров сходства их друг с другом. Это различие в вероятностях можно продемонстрировать и измерить, используя введенную в гл. 1 методику конкурирующих движений. Для сравнения значений сродства определенный элемент, например точка или отрезок, показывается поочередно с изображением пары других элементов той же длины, ориентации и удаленности, причем один из них находится слева от первого элемента, а другой — справа. В результате будут восприниматься два одновременных движения: обычно исходная линия расщепляется и участвует в движении сразу в *двух* направлениях (рис. 2.3а). Заметим, что в этом при-

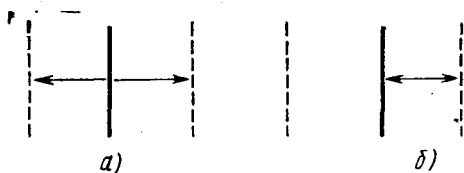


Рис. 2.3. Влияние расстояния на степень сродства между двумя отрезками

мере функция установления соответствия не является взаимно-однозначным отображением, поскольку исходный отрезок связывается более чем с одним элементом. Если теперь одно из расстояний возрастает, то вероятность увидеть только одно движение (по направлению к ближайшему соседу) также возрастает вплоть до достижения той точки, где уже воспринимается только одно это движение (рис. 2.3б). Сопоставления, таким образом, могут отличаться по силе результатов их восприятия (перцептов), которая может быть измерена по относительной частоте их возникновения в восприятии. Надо также отметить, что восприятие расщепляющегося движения не является единственным перцептом, связанным с этой конфигурацией из равноудаленных элементов (хотя такое движение, несомненно, является наиболее вероятным при условии, что точка фиксации взгляда удерживается в центре кадра). Одно из направлений движения может быть более предпочтительно, особенно тогда, когда наблюдение ведется достаточно долго, а предпочитаемое направление движения со временем может изменяться. Если, однако, сделать усреднение по времени

(или по наблюдателям), то эти два направления будут примерно равновероятны<sup>2.6</sup>.

Предпочтение, отдаваемое одному сопоставлению перед другим, зависит не только от расстояний, но также и от других параметров, таких как ориентация, длина и яркость (или контраст). Эти предпочтения указывают на то, что в зрительной системе имеется мера сродства, направляющая установление соответствия между изолированными опорными символами. В рамках схемы установления соответствия, использующей сопоставление элементарных опорных символов, можно ожидать, что такое выражение сродства является простейшей операцией, на которой базируется процесс сопоставления. В этом смысле сродство является «встроенной» метрикой: вычисление его представляет собой атомарную операцию, на которой основан процесс установления соответствия.

В данном разделе функция сродства будет изучена с помощью сравнения степеней сродства у различных пар элементов. Вначале будут исследованы изолированные точки, за ними — отрезки различной длины и ориентации<sup>2.7</sup>. Будет показано, что для изолированных точек на сродство оказывают влияние расстояния между точками и их яркость. Исследуются три условия, которые оказывают влияние на степень сродства и потому должны приниматься во внимание при сравнении степеней сродства: «периферийный прирост», «периферийное выравнивание» и «выравнивание за счет МСИ».

Непосредственный метод сравнения степеней сродства заключается в следующем. Элемент  $X$  показывается в конкурирующем движении с двумя другими элементами:  $Y1$  справа от него;  $Y2$  — слева. При этом отмечается, какое из двух возможных движений является более сильным или более вероятным. Похожий на этот, но иногда более полезный метод основан на том факте, что функция сродства монотонно убывает с увеличением расстояния. Если расстояние между  $X$  и  $Y2$  возрастает, то убывает вероятность увидеть  $X$  в движении с  $Y2$  до того момента, когда при некотором значении расстояния эта вероятность не станет равна нулю и будет восприниматься только движение к  $Y1$ . Затем это «нулевое расстояние» сравнивается для разных  $Y2$ <sup>2.8</sup>. Стоит, наверное, отметить, что методика конкурирующих движений не позволяет непосредственно измерить изучаемые степени сродства. Как будет обсуждаться в следующем разделе, в ситуации конкурирующих движений из-за конкурирующих взаимодействий степени сродства изменяются. Однако при разумных допущениях направление, предпочитаемое в условиях конкурирующих движений, может служить указателем более высокой степени сродства<sup>2.9</sup>.

Сперва рассмотрим простую ситуацию, когда единственная точка  $X$  предъявляется в конкурирующем движении с двумя другими точками: с одной стороны,  $Y1$  на расстоянии  $d_1$ , с другой —  $Y2$  на расстоянии  $d_2$  (рис. 2.4a; наблюдатель фиксирует свой взгляд на точке  $X$ ). Если  $d_1 = d_2$  и точки имеют одинаковую яркость, то  $X$ , как правило, расщепляется и движется одновременно

в обоих направлениях. Если  $d_1 < d_2$ , то преобладает движение в сторону  $Y1$ , в то время как при  $d_2 < d_1$  предпочтительным оказывается движение в сторону  $Y2$ , и это предпочтение становится все сильнее по мере увеличения  $d_2$ . При расстояниях  $d_1 = 0,5^\circ$  зрительного угла и  $d_2 = 1,5^\circ$  (при времени предъявления 120 мс и МСИ 40 мс) предпочтение практически всегда отдавалось  $Y1$  (рис. 2.4б).

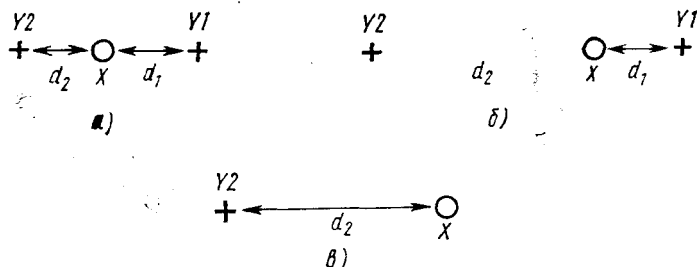


Рис. 2.4. Влияние расстояния на степень сродства между точками: а — расщепленное движение; б — предпочтение отдается  $Y1$ ; в — если  $Y1$  удален, то X и  $Y2$  видны в гладком движении

Экспериментальные изображения на конкурирующие движения вскрывают одну из причин того, почему оценки «гладкости» движения не подходят для изучения проблемы установления соответствия. Рассмотрим, к примеру, такую конфигурацию, в которой  $Y2$  находится на «нулевом» расстоянии от X, а затем удалим отсюда  $Y1$ . Тогда X немедленно станет виден в гладком движении с  $Y2$  (рис. 2.4в). Это демонстрирует крайнюю нечувствительность гладкости в отношении уровня сродства. Хотя сродство  $X \leftrightarrow Y1$  явно больше сродства  $X \leftrightarrow Y2$ , но обе пары, если их предъявлять по отдельности, видны в одинаково гладком движении. В разд. 2.4 мы изучим вторую фундаментальную проблему оценок гладкости и придем к выводу, что гладкость нельзя принять за меру сродства.

Второй после расстояния параметр, влияющий на сродство, — это относительная яркость. Его влияние не очень сильно, пока точки не станут значительно отличаться по своей яркости; тогда сродство будет больше между точками с близкими значениями яркости.

Первый и несомненный вывод из вышеописанных опытов таков: в противоположность утверждениям предшествовавших авторов пары опорных элементов все-таки отличаются друг от друга своим сродством. Второй вывод тот, что сродство у пар точек зависит и от расстояния между ними, и от их относительной яркости.

Сродство изменяется также в зависимости от положения изображения в зрительном поле, т. е. по отношению к сродству зрительное поле не вполне однородно. Периферийная часть поля отличается от центра в двух отношениях.

1. *Периферийный прирост.* Степень сродства возрастает от центра к периферии. Это усиление означает, что точка фиксации взгляда может оказывать ощутимое влияние на воспринимаемое соответствие. Если описанная выше конфигурация, иллюстрирующая конкурирующие движения (рис. 2.4а), рассматривается при фиксации взгляда на  $Y1$ , то предпочтительным оказывается движение  $X \rightarrow Y2$ , так как в этом случае  $Y2$  расположен ближе к периферии, чем  $Y1$ . Наоборот, при фиксации на  $Y2$  возрастает степень сродства  $X \rightarrow Y1$  и преобладает движение с  $Y1$ .

2. *Периферийное выравнивание.* Вторым результатом смещения изображения в сторону периферии — это меньшая выраженность воспринимаемых предпочтений, возникающих в результате различий в расстоянии. Иными словами, при движении к периферии степени сродства стремятся выровняться. Впрочем, предпочтение сохраняет свой знак, т. е. направление предпочтения не изменяется.

Еще одно условие, влияющее на степени сродства, связано с временными характеристиками предъявляемых изображений, в частности с межстимульным интервалом.

3. *Выравнивание за счет МСИ.* Возрастание МСИ производит эффект, аналогичный периферийному выравниванию, т. е. предпочтение становится менее ощутимым, но не меняет своего знака. Увеличение времени предъявления первого кадра оказывает действие, сходное с эффектом увеличения МСИ<sup>2.10</sup>. (Количественные измерения выравнивания за счет МСИ для случая отрезков приведены в работе [Ульман, 1978 а], рис. 1.)

Перечисленные выше эффекты влияния местоположения и временных интервалов могут быть вполне ощутимы. Значит, при проведении экспериментов по изучению сродства надо обязательно обращать особое внимание на точку фиксации взгляда и на используемый МСИ.

Теперь вернемся к сравнению сродства отдельных коротких отрезков.

**Сродство отрезков.** Те же параметры, что и для точек, оказываются важными в случае отрезков, но для отрезков возникает ряд новых параметров. Вначале рассмотрим влияние на степень сродства отрезков расстояния между ними, а затем изучим влияние их длины и ориентации<sup>2.11</sup>.

Влияние расстояния. Сродство отрезков, как и сродство отдельных точек, зависит от их пространственной близости. Эта зависимость иллюстрируется рис. 2.5. Одиночный отрезок  $s$  испытывается в конкурирующем движении с двумя отрезками:  $l$  — слева от него и  $r$  — справа. Эти два боковых отрезка параллельны  $s$  и имеют ту же длину, но  $l$  ближе к  $s$ , чем  $r$ . Как уже отмечалось, предпочтительным является движение в сторону  $l$ , и по мере роста расстояния между  $r$  и  $s$  это предпочтение становится все значительнее.

Влияние пространственной близости труднее оценить в том случае, когда используются непараллельные отрезки. Что, к при-

меру, считать расстоянием между отрезками  $c$  и  $r$  на рис. 2.6? Расстояние между их центрами, между их концами, среднее расстояние или, может быть, какой-то другой параметр? Выбор конкретного способа определения расстояния является предварительным условием для оценки влияния других параметров. Предположим, к примеру, что на рис. 2.7 движение в сторону  $l$  оказы-

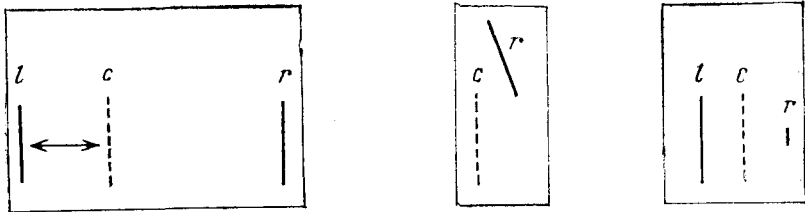


Рис. 2.5.—2.7. С помощью критерия отбора расстояние между отрезками определено как расстояние между их центрами

вается предпочтительнее движения в сторону более короткого отрезка  $r$ . Если в качестве эффективного расстояния рассматривать расстояние между центрами отрезков, то это предпочтение можно отнести на счет разности длин отрезков, поскольку  $r$  и  $l$  равноудалены от  $c$ . Если расстояние определяется по-другому, например между концами отрезков, то это предпочтение можно отнести, по крайней мере отчасти, на счет того факта, что теперь расстояние между  $c$  и  $r$  больше расстояния между  $c$  и  $l$ . Следовательно, зависимость сродства от такого параметра, как длина, будет иметь разную значимость. Главное соображение при выборе конкретной метрики заключается в том, что она должна приводить к простому описанию функции сродства. Желательно, например, чтобы для параллельных линий одинаковых длин и яркости сродство можно было выразить как функцию одного только расстояния. Этот критерий выбора может быть, по крайней мере приблизительно, удовлетворен, если расстояние между отрезками определить как расстояние между их центрами. Если сравнивать расстояния между серединами отрезков, то на рис. 2.8 можно ожидать, что степени сродства окажутся равными, и наблюдаемое в эксперименте движение оправдывает это ожидание.

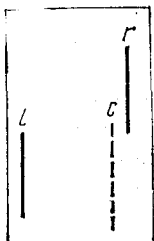


Рис. 2.8.

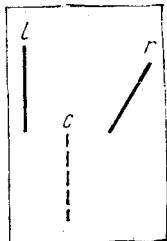


Рис. 2.9.

Рис. 2.8. При измерении расстояния между центрами отрезков отрезки  $l$  и  $r$  равноудалены от  $c$

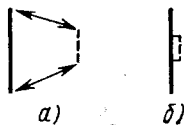
Рис. 2.9. В этой конфигурации  $l$  и  $r$  примерно равноудалены от  $c$  с точки зрения различных упомянутых в тексте метрик

Если же, к примеру, используется метрика, основанная на минимальном расстоянии, то это равенство нельзя объяснить на основе одних расстояний. Следовательно, метрика, основанная на минимальном расстоянии, в силу нашего критерия отбора оказывается неподходящей. Подобным же образом отвергаются и другие метрики, тогда как расстояние между центрами, по-видимому, действительно удовлетворяет указанному критерию в большинстве случаев и потому будет нами принято. Кроме того, большинство используемых во всем этом разделе конфигураций выбрано так, чтобы свести к минимуму различия между разными способами определения расстояния (рис. 2.9).

**Влияние ориентации.** Сродство между отрезками является убывающей функцией различия их ориентаций. На рис. 2.9 приведен пример, в котором отрезок  $s$  испытывается в конкурирующем движении с отрезками разной ориентации. Отрезок  $l$  слева имеет ту же ориентацию, что и  $s$ , а  $r$  — другую. Два конкурирующих отрезка имеют одинаковые длину и яркость, и расстояния их до  $s$  тоже одинаковы. В таких условиях предпочтительным оказывается движение в сторону  $l$ , и это предпочтение возрастает с ростом угла между  $s$  и  $r$ .

**Влияние длины.** Сходство по длине усиливает сродство отрезков. При прочих равных условиях отрезок в конкурирующем движении будет предпочтительнее сопоставляться с тем отрезком, который ближе к нему по длине. Если сравниваются отрезки разной длины, то соответствие обычно устанавливается между их *концевыми точками*<sup>2,12</sup>. В этом случае можно наблюдать, как одиночный отрезок движется, меняя свою длину (рис. 2.10а). Если, однако, расстояние между отрезками достаточно мало, то установление соответствия между их концами происходить не будет. Тогда более короткий отрезок виден в движении только с *частью* более длинного отрезка (рис. 2.10б).

Рис. 2.10. Установление соответствия между концевыми точками (а), между отрезком и подотрезком (б)



Сравнение влияний расстояния, ориентации и длины. С использованием различных конфигураций в экспериментах на конкурирующее движение были установлены относительные влияния на степень сродства пространственной близости, ориентации и длины отрезков.

Результаты сведены в табл. 2.1, где даны значения относительной силы разных влияний. Например, из этой таблицы можно видеть, что результат влияния различия ориентаций в  $60^\circ$  сравним с влиянием отношения расстояний, равного 2,25, или отношения длин, равного 2,1. Всем трем указанным значениям соответствует сила влияния примерно в 3 единицы на шкале от 0 (нижнего предпочтения) до 4 (абсолютное предпочтение).

Таблица 2.1. Относительное влияние различия ориентаций, отношения расстояний и отношения длин

	0	1	2	3	4
Различие ориентаций (в градусах)	15	30	45	60	75 90
Отношение расстояний	1,1	1,2	1,6	2,25	2,7 3,8
Отношение длин	1,04	1,13	1,5	2,1	2,5
$1/\cos \alpha$	1,04	1,15	1,41	2,0	—

Условия предъявления. Длина центральной линии составляла около  $1^\circ$  зрительного угла, время предъявления — 120 мс, МСИ — 40 мс; взгляд испытуемого фиксировался на центре среднего отрезка. Параметр длины выражен через отношение, где одна из линий имеет размер в  $1^\circ$  зрительного угла. Таким образом, отношение длин, равное 2, означает, что один из боковых отрезков имеет длину только  $0,5^\circ$ . Расстояния выражены так же: один из отрезков был зафиксирован на расстоянии  $1^\circ$  от центральной линии, а расстояние до второго отрезка подвергалось изменению. Интересно отметить, что в данных экспериментальных условиях абсолютные значения длины не играли почти никакой роли, так как действительно определяющим параметром было отношение длин. Если, например, наклон в  $30^\circ$  компенсирует отношение расстояний, равное 1,2, то это отношение между ними сохраняется в широком диапазоне длин отрезков<sup>2,13</sup>. Эти данные усреднены по двум наблюдателям, каждый из которых давал десять оценок каждой конфигурации. Надо отметить, что данные результаты были получены для значений времени предъявления 100 мс и более, при положении изображения в центре зрительного поля. Имеются некоторые указания на то, что влияние ориентации и длины может зависеть от времени предъявления, равно как и от положения изображения в зрительном поле.

Примеры использования данных таблицы:  $X$  будет обозначать центральный отрезок на первом изображении (он всегда вертикален),  $Y1$  и  $Y2$  составляют второе изображение.

а.  $Y2$  отклоняется на  $60^\circ$  от вертикали,  $Y1$  вертикален. Отрезку  $Y1$  оказывается предпочтение силой в 3 единицы — это «значительное предпочтение».

б. Отрезок  $Y1$  наклонен под углом  $45^\circ$ , и в то же время длина отрезка  $X$  в 1,5 раза больше длины  $Y2$ . В таблице показано, что в этой ситуации никому не отдается предпочтение и оба движения будут равновероятны (рис. 2.11).

Как можно видеть из таблицы, сродство отрезков связано с их сходством. Следующий интересный вопрос — связь между относительными влияниями различия в ориентации отрезков и отношением их длин. Рассмотрим две проекции отрезка: сначала он параллелен плоскости изображения, а затем повернут на  $\alpha^\circ$ . Если

вращение происходит в плоскости изображения, то получившиеся два отрезка (до и после вращения) имеют одинаковую длину, но отличаются по направлению на  $\alpha^\circ$ . Если отрезок отклоняется *в глубину*, то он не изменяет своего направления, а лишь уменьшается в длине\*. Отношение начальной длины к конечной (пре-

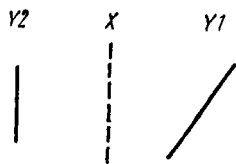


Рис. 2.11. Согласно табл. 2.1 влияние разности направлений в  $45^\circ$  эквивалентно влиянию отношения длин, равного 1,5

небрегаем перспективными эффектами) равно  $1/\cos \alpha$ . Если при начальной ориентации отрезок не параллелен плоскости изображения, то при достаточно малом угле вращения  $\alpha^{**}$  сохраняется та же связь. В табл. 2.1 показано, что влияние разности направлений в  $\alpha^\circ$  и влияние отношения длин, равного  $1/\cos \alpha$ , сопоставимы между собой (вплоть до примерно  $75^\circ$ ). Таким образом, сродство отражает различие направлений в пространстве. Отсюда можно ожидать, что на изображении объекта, вращающегося в пространстве, распределение соответствий будет одинаковым при вращении объекта в плоскости изображения или перпендикулярно ей. В этом вопросе могут оказаться полезными эксперименты Шепарда и Джадда (1976) с вращающимися фигурами. Эти исследователи обнаружили, что прекращение установления согласованного соответствия между изображениями вращающейся фигуры происходит при одних и тех же углах поворота и в глубину, и в плоскости изображения.

### 2.3. Взаимодействия более высокого уровня

#### 2.3.1. Конкуренция при расщеплении и при слиянии

При изучении сложных экспериментальных изображений, состоящих из большего числа опорных символов, распределение соответствий уже нельзя объяснить на основе одного лишь сродства. Как будет показано, взаимодействия между элементами изменяют исходные степени сродства, так что элемент, сопоставляемый данному элементу, уже не обязательно является элементом с наибольшей степенью сродства. Таким образом, роль функции сродства аналогична роли основных функций  $P_i$  на рис. 2.2б, а взаимодействия более высокого уровня, которые мы собираемся изучать, играют при этом роль функций  $f_i$ . Для различения исходных степеней сродства и окончательного распределения соответствий введем понятие степени соответствия, сокращенно  $CS$

\* С точки зрения наблюдателя. — Прим. ред.

\*\* На самом деле это верно лишь при малых отклонениях от плоскости изображения исходного отрезка. — Прим. ред.



(от «correspondence strength»). Если на сложном изображении сопоставление  $x$  с  $y$  предпочтительнее, чем сопоставление  $x$  с  $z$ , то делаем вывод, что  $CS(x, y) > CS(x, z)$ , т. е.  $CS$  между  $x$  и  $y$  больше, чем  $CS$  между  $x$  и  $z$ . Если  $CS(x, z) = 0$ , то  $x$  и  $z$  никогда не будут видны во взаимном движении<sup>2,14</sup>.

Теперь задачу данного раздела — провести исследование установления соответствия в сложных конфигурациях — можно сформулировать на языке степени соответствия: нужно найти метод вычисления  $CS$  по исходным степеням сродства. Главный результат состоит в том, что  $CS$  получается из степеней сродства опорных символов с помощью простых локальных взаимодействий. Значимость этого вывода имеет двоякий характер. Во-первых, он подтверждает результат гл. 1 о том, что установление соответствия в общем случае основано на сопоставлении элементарных опорных символов. Во-вторых, он будет играть важную роль в развитии вычислительной теории установления соответствия в гл. 3. В настоящем разделе будут обсуждаться два типа взаимодействий, которые назовем *конкуренцией при расщеплении* и *конкуренцией при слиянии*. В следующем разделе результаты, касающиеся сродства и взаимодействий, будут объединены в одно обобщение, называемое *схемой состязаний*. В разд. 2.4 схема состязаний будет применена к нескольким примерам, чтобы показать, что установление соответствия действительно согласуется со схемой состязаний.

При конкурирующем движении попеременно показываются один элемент  $X$  и два других элемента  $Y1$  и  $Y2$ . Получающееся в результате возвратно-поступательное движение элементов состоит из двух фаз — расщепления и слияния, которые теперь рассмотрим в отдельности. В фазе расщепления один элемент  $X$  заменяется парой элементов  $Y1$  и  $Y2$ . В режиме слияния сначала предьявляется пара элементов, а затем — один элемент. Каждая из фаз может быть изучена сама по себе с применением тахистоскопического способа предьявления, т. е. такого способа, когда в быстрой последовательности предьявляются два кадра с последующим интервалом между сериями в несколько секунд. Если сначала показывается  $X$ , а затем —  $Y1$  и  $Y2$ , то наблюдается только фаза расщепления. Для исследования фазы слияния порядок предьявления меняется на обратный: предьявляются  $Y1$  и  $Y2$ , а затем один  $X$ .

*Конкуренция при расщеплении.* Мы уже видели, что в ситуации конкурирующих движений (см. рис. 2.3 и 2.4) движение между  $X$  и  $Y2$  прекращается при наличии другого элемента  $Y1$ , достаточно близкого к  $X$ . То же самое явление имеет место и для каждой из фаз в отдельности. Если после  $X$  предьявляется один  $Y2$ , то видно, как  $X$  сопоставляется с  $Y2$  и участвует в движении с ним. В присутствии же  $Y1$  наблюдается только движение  $X \rightarrow Y1$ . Это прекращение движения означает, что вследствие конкуренции окончательное значение  $CS(X, Y2)$  меньше исходного значения сродства  $AF(X, Y2)$ . Эта ситуация, по-видимому, симмет-

рична<sup>2,15</sup>, т. е.  $CS(X, Y1) < AF(X, Y1)$ , что иллюстрируется схемой на рис. 2.12. Окончательное значение  $CS$  между  $X$  и  $Y1$  зависит и от  $AF(X, Y1)$ , и от  $CS(X, Y2)$ ; то же верно и для значения  $CS$  между  $X$  и  $Y2$ . В общем случае степень соответствия  $CS(A, B)$  между двумя данными элементами  $A$  и  $B$  уменьшается за счет конкуренции при расщеплении с любым другим элементом  $Z$ , для которого  $CS(A, Z)$  не равно нулю.

**Конкуренция при слиянии.** Если фаза слияния в конкурирующем движении изолирована, то наблюдается тот же самый эффект: движение  $X \leftarrow Y2$  прекращается в присутствии элемента  $Y1$ . Эффекты расщепления и слияния, по-видимому, имеют равную силу, а именно влияние  $Y1$  на воспринимаемое движение между  $X$  и  $Y2$  примерно одинаково в обеих фазах. В качестве примера конкуренции при слиянии в несколько более сложной ситуации рассмотрим рис. 2.13. На рис. 2.13а  $X1$  виден расщепляющимся и движущимся в обоих направлениях, тогда как на рис. 2.13б  $X1$  участвует только в движении направо. Все происходит так, словно  $X2$  (который может находиться с любой стороны от  $Y2$ ) удерживает  $Y2$  и тем самым предотвращает установление соответствия  $X1 \rightarrow Y2$ . Это означает, что  $AF(X2, Y2)$  уменьшает значение  $CS(X1, Y2)$ , т. е.  $CS(X1, Y2) < AF(X1, Y2)$  в результате взаимодействия.

Эти два типа состязаний иллюстрируются схемой на рис. 2.14. Сплошные линии показывают ненулевое сродство двух элемен-

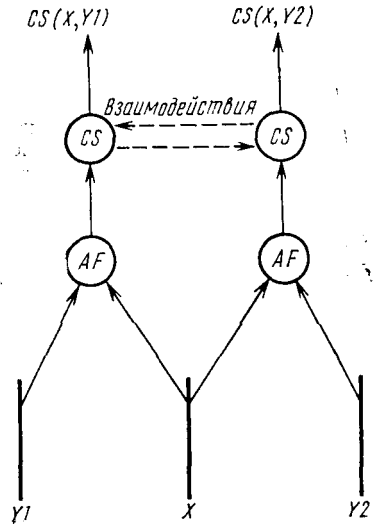


Рис. 2.12. Значения  $CS$  получаются из исходных значений сродства в результате локальных взаимодействий

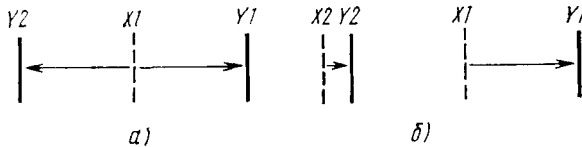


Рис. 2.13. Конкуренция при слиянии:  $X1$  расщепляется (а),  $X1$  не расщепляется благодаря конкуренции со стороны  $X2$  (б)

тов. Пунктирные линии обозначают взаимодействия конкуренций: горизонтальные представляют конкуренцию при расщеплении, вертикальные — при слиянии. Изображение  $A1$  может сопоставиться с  $B1, B2$  или  $B3$ , следовательно, они все взаимодействуют через конкуренцию при расщеплении. Подобная же ситуация со-

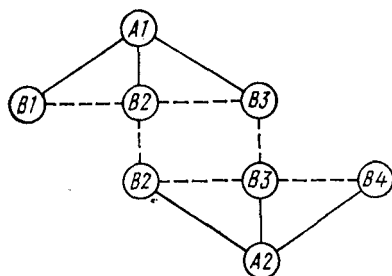


Рис. 2.14. Диаграмма, объединяющая конкуренцию при расщеплении и при слиянии

такой сети, где сплошные линии символизируют взаимоподдержку, пунктирные — конкуренцию. Мы еще вернемся к этому вопросу в гл. 3.

храняется для сопоставлений между A2 и B2, B3, B4. Изображения A1 и A2 имеют B2 и B3 возможными парными элементами при сопоставлении, следовательно, эти пары взаимодействуют между собой через конкуренцию при слиянии. Хотя при современном состоянии знаний возникают непреодолимые трудности в случае попытки связать описанные выше функции с известными физиологическими структурами, рис. 2.14 наводит на мысль о возможной общей схеме

### 2.3.2. Схема состязаний

Изучение простых экспериментальных изображений позволило выявить несколько принципиальных положений, лежащих в основе процесса установления соответствия. Назовем схему, объединяющую все описанные до сих пор принципы, *схемой состязаний* для вычисления соответствия. Схема состязаний определена не полностью, поскольку еще неизвестны точные выражения для средства и функций взаимодействия. Для большей ясности принципы, включенные в схему состязаний, конспективно перечислены ниже.

1. *Сродство*. Для пар опорных элементов может быть указана мера сродства, которая зависит прежде всего от расстояния между ними, различных параметров сходства, временных промежутков и условий рассмотрения изображений. Параметры сходства включают в себя яркость, ориентацию и длину. Влияния временных промежутков и условий рассмотрения сведены в следующие три пункта.

2. *Периферийный прирост*. Степень сродства возрастает от центра к периферии зрительного поля.

3. *Периферийное выравнивание*. Различия между степенями сродства менее заметны на периферии по сравнению с центром зрительного поля.

4. *Выравнивание за счет МСИ*. Влияние увеличения МСИ заключается в выравнивании степеней сродства и тем самым маскировании предпочтений при установлении соответствия. Окончательное значение степени соответствия получается из исходных степеней сродства с помощью двух типов конкурентных отношений, которые сформулированы ниже.

5. *Конкуренция при расщеплении*. Если данный опорный символ имеет несколько возможных парных символов при сопостав-

лении, то все те из них, которые имеют ненулевую степень соответствия, конкурируют друг с другом.

6. *Конкуренция при слиянии.* Если несколько опорных символов имеют общий возможный парный элемент при сопоставлении, то все те из них, которые имеют ненулевую  $CS$ , конкурируют друг с другом.

В следующем разделе схема состязаний будет проверена применительно к конфигурациям, более сложным, чем те, на основе изучения которых она была построена.

## 2.4. Примеры применения схемы состязаний

В предыдущем разделе с помощью комбинации нескольких правил, направляющих процесс установления соответствия на простых экспериментальных изображениях, была получена схема состязаний для установления соответствия. Возможно, конечно, что в определении процесса сопоставления участвуют и другие правила. С методологической точки зрения, однако, целесообразно занять консервативную позицию и воздержаться от введения новых правил (если только схема состязаний не окажется явно недостаточной), избегая тем самым путаницы, которая в противном случае может возникнуть от разрастания множества необоснованных правил.

В этом разделе схема состязаний используется в нескольких примерах с целью проверки ее достаточности в более сложных ситуациях. Показано, что воспринимаемое в действительности соответствие согласуется со схемой состязаний даже в тех некоторых случаях, которые раньше рассматривались как противоречащие теориям восприятия движения (Колерс, 1972; Эттнев, 1974). Наша цель — показать, что эти, на первый взгляд затруднительные, случаи согласуются со схемой установления соответствия с помощью состязаний и могут быть предсказаны ею. Четыре случая будут изучены при сравнении воспринимаемого соответствия с результатами, предсказанными конкретной версией схемы состязаний (другие примеры можно найти в диссертации [Ульман, 1977а]). Под «использованием конкретной версии» мы понимаем некоторый конкретный выбор функций сродства и функций состязания. Так как для определения действительной формы этих функций не достаточно имеющихся данных, то детальное экспериментирование над конкретной моделью является преждевременным. Дело, однако, заключается в том, чтобы показать, что эти примеры могут быть объяснены схемой состязаний и, следовательно, не требуют никаких дополнительных схемы взаимодействия. Схема состязаний, используемая в этом разделе, подробно описана в работе [Ульман, 1977а, приложение 4].

### 2.4.1. Четыре примера

В каждом примере два кадра показаны в условиях кажущегося движения. На каждом кадре представлен либо единственный элемент (маленькое пятно размером 12 мин зрительного угла), либо несколько элементов, расположенных вдоль горизонтальной строки<sup>2,16</sup>. Для удобства пояснений второе предъявление нарисовано под первым, хотя на самом деле они оба располагались на одной линии. Элементы первого предъявления обозначены  $A1, A2$  и т. д., второго —  $B1, B2$  и т. д. Числа, помечающие связи между элементами, представляют собой

значения сродства или *CS* между двумя элементами. Каждый пример сопровождается тремя рисунками. На первом представлены исходные значения сродства, использованные при вычислении. Эти значения [см. работу Ульмана, 1977а, приложение 4] были выбраны согласующимися с правилами схемы состязаний. Применены два типа функции сродства. Первая, называемая условием соответствия по соседнему элементу, используется в экспериментах с малым МСИ и при взгляде, сосредоточенном в центре поля зрения. Вторая, называемая условием соответствия «по связанному элементу» (см. ниже), получается из первой в результате выравнивания степеней сродства. Второй рисунок каждого примера представляет окончательные значения степени соответствия, получающиеся в результате совместного влияния степеней сродства и конкурирующих взаимодействий. Третий рисунок иллюстрирует соответствие, наблюдавшееся в действительности. Поскольку некоторые варианты сопоставления наблюдаются чаще других, то принято следующее обозначение. Сплошные линии обозначают надежное, устойчивое сопоставление, а пунктирные — соответствие более слабое и менее устойчивое. Большинство из этих экспериментальных изображений использует конфигурации, примененные в экспериментах Колерса (1972)<sup>2.17</sup>.

**Пример 1: тройкое расщепление.** Расщепляющееся движение не ограничивается только парами элементов. Колерсом (1972) были исследованы тройкие расщепления, пример которых приведен на рис. 2.15. В колюм соответствии с данными реального восприятия схема состязаний свидетельствует о предпочтении, отдаваемом ближайшему соседу, которое при больших МСИ становится меньше. При наблюдении с малым МСИ *B0* и *B2* (правый и левый элементы) кажутся неподвижными. При больших МСИ все три элемента видны в движении, однако движение между *A* и *B1* остается самым сильным.

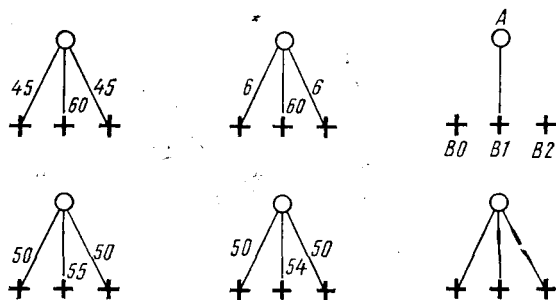
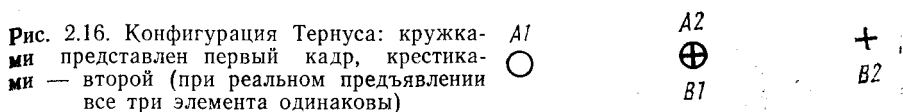


Рис. 2.15. Тройкое расщепление: при малом МСИ большое предпочтение отдается самому ближайшему элементу, при большем МСИ наблюдается расщепление

**Пример 2: «конфигурация Тернуса».** Этот пример разберем несколько подробнее, поскольку он позволяет выявить интересную зависимость восприятия от используемой величины МСИ и поскольку он служил в качестве аргумента против того, чтобы считать процесс установления соответствия основанным на опорных элементах [Тернус, 1926]. В этом примере каждый кадр экспериментального изображения содержит два элемента: элементы *A1* и *A2* на первом кадре, элементы *B1* и *B2* на втором<sup>2.18</sup>. Элементами могут быть либо точки, либо параллельные отрезки, причем, как показано на рис. 2.16, положения *A2* и *B1* совпадают. Соответствия воспринимаются одним из двух способов, что

зависит от таких параметров экспериментального изображения, как расстояния и МСИ. При первом способе  $A1$  связывается с  $B1$ , а  $A2$  — с  $B2$ , т. е. пара элементов движется направо как нечто целое. Назовем этот способ восприятия движения *способом установления соответствия по связанному элементу*. При втором способе устанавливается соответствие  $A2 \rightarrow B1$  (т. е.  $A2$  связывается в



пару с  $B1$ ), а  $A1$ , если он вообще виден в движении, связывается с  $B2$ . В этом случае элемент  $A2$  связывается со своим ближайшим соседом, и поэтому назовем этот тип воспринимаемого соответствия *способом установления соответствия по соседнему элементу*. В гештальтпсихологии [Тернус, 1926; Коффка, 1935] способ по связанному элементу, при котором  $A2$  не связывается в пару со своим ближайшим соседом  $B1$ , истолковывался как свидетельство того, что соответствующий  $A2$  элемент на втором кадре определяется его «ролью в поле». Это в неявной форме означает, что не изолированные элементы, а «поля» в целом являются опорными символами соответствия<sup>2.19</sup>. Но наблюдаемое в эксперименте соответствие согласуется на самом деле с намеченной выше схемой состязаний. Хотя  $B1$  и является ближайшим соседом  $A2$ , конкурирующие взаимодействия все же могут привести к тому, что в конечном счете сопоставления будут такими:  $A1 \rightarrow B1$ ,  $A2 \rightarrow B2$ . Чтобы понять, почему это происходит, вспомним, что окончательное значение степени соответствия между двумя элементами зависит как от начальной степени родства между ними, так и от взаимодействий, которые уменьшают степень соответствия. Таким образом, хотя степень родства между  $A2$  и  $B1$ , конечно, больше степени родства  $A2 \rightarrow B2$ , конкурирующие между собой взаимодействия могут сделать окончательное значение  $CS$  на паре  $A2 \rightarrow B1$  меньше значения  $CS$  между  $A2$  и  $B2$ . Вопрос о том, какой способ установления соответствия будет восприниматься при данных экспериментальных условиях, нельзя решить без учета как степеней родства, так и участвующих при этом взаимодействий. Можно, однако, дать качественное предсказание. Например, можно ожидать, что при фиксированном МСИ способ установления соответствия по связанному элементу будет преобладать при малых расстояниях между элементами, а при больших расстояниях — способ по соседнему элементу, что и наблюдается в действительности<sup>2.20</sup>.

Описанная схема состязаний, построенная с учетом взаимодействий, выделенных в простых случаях, достаточна для объяснения (по крайней мере, качественного) установления соответствия, наблюдаемого и в конфигурации Тернуса. Это объяснение является более полным, чем описание с помощью понятия «поле», поскольку оно предсказывает зависимость наблюдаемого способа установления соответствия от расстояния между элементами. С методологической точки зрения кажется нежелательным прибегать к введению каких-либо новых взаимодействий или правил организации до тех пор, пока не будет показано, что изложенных выше двух типов конкурирующих взаимодействий оказывается недостаточно.

**Влияние МСИ на конфигурацию Тернуса.** Разные факторы в состоянии вызывать переход от одного способа восприятия соответствия в конфигурации Тернуса ко второму. Например, при диоптрическом предъявлении (когда один кадр

проецируется на один глаз, а второй — на другой) преобладает способ по связанному элементу [Пэнтл и Пиччиано, 1976]. Среди других параметров, влияющих на способ восприятия, — расстояния между элементами, относительная интенсивность различных элементов и МСИ. Влияние МСИ заключается в том, что при малом МСИ преобладает способ восприятия конфигурации Тернуса по соседнему элементу, а при большом МСИ верх берет способ по связанному элементу. Плавное увеличение МСИ вызовет довольно резкий переход от одного способа к другому [Пэнтл и Пиччиано, 1976].

Такой эффект, обусловленный ростом МСИ, возникает не только в случае двухточечной конфигурации Тернуса. Пэнтл и Пиччиано (1976), используя трехэлементное экспериментальное изображение, наблюдали аналогичный переход от одного способа восприятия к другому. Такие же способы восприятия можно наблюдать на двухкадровых экспериментальных изображениях, состоящих из случайно распределенных точек. Первый кадр содержит точки, разбросанные по экрану случайным образом. Второй кадр состоит из того же набора точек, сдвинутых как единое целое на некоторую величину. При малом МСИ намечается тенденция к наблюдению локальных, несвязанных движений. При большом МСИ преимущественным становится способ восприятия по связанному элементу, т. е. возрастает тенденция увидеть одинаковый параллельный перенос всего набора точек. Пэнтл и Пиччиано, изучавшие явление смены способов восприятия, пришли к выводу, что смена происходит при МСИ, примерно равном 40 мс. Этот результат, однако, не универсален в том смысле, что у разных конфигураций смена способов восприятия происходит при разных МСИ. В конфигурации Тернуса, например, чем больше расстояние между элементами, тем больший МСИ требуется для этой смены.

**Возможное объяснение: выравнивание за счет МСИ.** Переход к способу восприятия по связанному элементу, происходящий при возрастании МСИ, может быть предсказан на основании нашей схемы состязаний. При этом вовсе не надо предполагать наличие дополнительных механизмов или существование двух отдельных систем [Пэнтл и Пиччиано, 1976]. Мы уже видели, что МСИ может оказывать влияние на процесс установления соответствия, приводя к выравниванию. Отмечалось, что степени родства между опорными элементами зависят от МСИ в том отношении, что при большем МСИ значения степеней родства становятся более одинаковыми. Как мы сейчас увидим, в рамках схемы состязаний явление выравнивания за счет МСИ может привести к переходу от способа восприятия по соседнему элементу к способу по связанному элементу. На рис. 2.17 результаты применения схемы состязаний к конфигурации

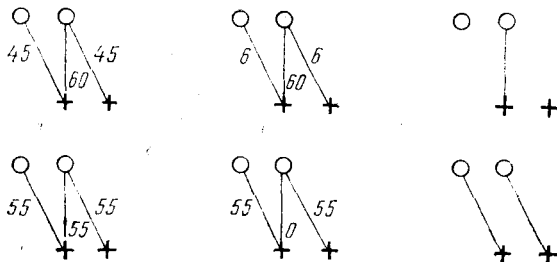


Рис. 2.17. Конфигурация Тернуса при малом (верхний ряд) и большом (нижний ряд) МСИ

Тернуса при малом МСИ сравниваются с частным случаем условий возникновения способа по связанному элементу — случаем, когда все степени сродства равны между собой. При малом МСИ самым сильным является сродство  $B \rightarrow V$ . Благодаря влиянию состязаний оно ослабляет и практически уничтожает связи  $A \rightarrow V$  (конкуренция при слиянии) и  $B \rightarrow C$  (конкуренция при расщеплении). Малый МСИ не оказывает влияния на связь  $A \rightarrow C$  (на рисунке не показана), и она остается равной исходному значению сродства. При большом МСИ совместной конкуренции связей  $A \rightarrow V$  и  $B \rightarrow C$  достаточно для уничтожения связи  $B \rightarrow V$ , и вследствие этого преобладает способ восприятия по связанному элементу. Кроме того, схема состязаний проявляет такое же бистабильное поведение, какое характеризует восприятие конфигурации Тернуса. По мере того как возрастание МСИ постепенно выравнивает степени сродства, происходит довольно резкий переход от полного преобладания способа восприятия по соседнему элементу к полному преобладанию способа восприятия по связанному элементу.

Если проведенный выше анализ верен и возникновение восприятия по связанному элементу действительно обусловлено выравниванием за счет МСИ, то можно сделать следующее предсказание. Поскольку сдвиг экспериментального изображения в сторону периферии производит эффект, аналогичный увеличению МСИ, то показ конфигурации Тернуса вне центра зрительного поля должен привести к усилению способа восприятия по связанному элементу. Это предсказание можно проверить, сравнивая воспринимаемое испытуемым движение при фиксации взгляда на центре изображения с движением при фиксации на точке, находящейся на угловом расстоянии  $1,5^\circ$  от центра. Как и ожидалось, способ восприятия по связанному элементу усиливается во втором случае.

Рассмотренное выше объяснение в некоторых отношениях чрезмерно упрощено. Добавочные осложнения возникают в результате группировки на более высоком уровне, как это описано во второй части подразд. 2.4.2, и в результате эффекта с малой зоной действия, упомянутого в разд. 1.2. Некоторые дополнительные вопросы, связанные с конфигурацией Тернуса, будут изучены в гл. 3 в рамках теоретических рассуждений.

**Смена способов восприятия и саккадическое подавление.** Глаза постоянно участвуют в движениях разных видов, одно из которых называется *саккадическим*. Саккада представляет собой баллистическое движение глазного яблока. Саккады происходят в среднем около 2 раз/с, являются движениями быстрыми и обычно довольно большой амплитуды (в среднем около  $5^\circ$  для произвольных саккад) [Дэвсон, 1969; Дитчберн, 1973].

С саккадами связано любопытное явление, называемое *саккадическим подавлением*. Не удивительно, что в течение длительности самой саккады, когда изображение стремительно перемещается по сетчатке, острота зрения крайне ослабляется. Более удивителен тот факт, что потеря четкого зрения значительно дольше самой саккады: она начинается примерно за 40 мс до саккады и прекращается спустя примерно 50 мс [Маккэй, 1970; Дитчберн, 1973]. Это явление потери зрительных ощущений вследствие саккад и называется *саккадическим подавлением*. Оно проявляется в возрастании промежутка времени, в течение которого зрение ухудшается, с 25—40 до более чем 100 мс на одну саккаду. Иными словами, при частоте 4 саккады/с острота зрения ослаблена в течение примерно 50% всего времени.



Можно предположить, что само существование саккадического подавления является прискорбным следствием некоторых свойств глаза. Однако Маккэй (1970) показал, что саккадическое подавление имеет место даже *без каких бы то ни было* движений глазного яблока. Если же само изображение быстро сдвигается по сетчатке, то уменьшение остроты зрения происходит почти по такой же схеме, что и при настоящей саккаде. Все происходит так, словно зрительная система всякий раз, когда сталкивается с большим и быстрым смещением, подавляет зрительную информацию в течение промежутка времени примерно 100 мс<sup>2.21</sup>.

Проведенное выше обсуждение двух способов восприятия соответствия указывает на возможность того, что саккадическое подавление служит полезной цели — преодолеть трудность, возникающую при большом и быстром смещении изображения по сетчатке. Уже упоминалось, что при определении соответствия важную роль играет *расстояние* между элементами. Чем оно меньше, тем выше это сродство. Следовательно, элементы стремятся образовать пару, сопоставиться со своими ближайшими соседями. Ожидать, что такой способ установления соответствия по соседнему элементу будет полезен, можно лишь до тех пор, пока скорость движения достаточно мала. Если же, например, берется киноплёнка с изображением быстро движущихся элементов, то нет никаких гарантий, что данный элемент на одном кадре можно корректно отождествить с его ближайшим соседом на следующем кадре. Из этого следует вывод, что всякий раз, когда смещение изображения по сетчатке между последовательными моментами съёмки велико, можно ожидать, что способ восприятия по соседнему элементу даст *случайные, неверные сопоставления*. Возможное решение этой проблемы — использование при таких условиях способа восприятия по связанному элементу, в котором ищется смещение, общее для всех элементов. Мы уже видели, что такое «переключение» от способа восприятия по соседнему элементу к способу по связанному элементу вызывается увеличением МСИ. Типичную саккаду можно представлять себе как два мгновенных снимка, разделённых МСИ, при котором зрение ослаблено и размыто и который длится около 25—40 мс. Этот временной интервал зачастую лежит в пределах, соответствующих способу восприятия по соседнему элементу. Возможным методом принудительного «переключения» на способ по связанному элементу является расширение интервала ухудшенного зрения за пределы длительности самой саккады. При обнаружении быстрых случайных смещений изображения также полезно осуществлять подобное подавление.

Наши знания о так называемом саккадическом подавлении все еще ограничены, и потому вышеприведенное предложение поневоле является умозрительным. Если оно, однако, сделано в правильном направлении и если саккадическое подавление есть проявление какого-то полезного процесса, то надо переключить наше внимание с поиска возможных связей между движениями глаза и ухудшенной остротой зрения на такие проблемы, как методы обнаружения быстрых смещений и методы управления остротой зрения.

**Пример 3: несимметричное движение.** Конфигурация, используемая в этом примере, показана на рис. 2.18. Хотя она и симметрична, воспринимаемое соответствие асимметрично: центральный элемент *A2* движется либо направо, либо налево, приводя к возникновению бистабильной конфигурации. Довольно редко соответствие принимает симметричный вид, и центральный элемент вообще не движется или, что бывает очень редко, он расщепляется и движется в двух на-

правлениях. В рамках схемы состязаний симметричное движение возникает тогда, когда исходные степени сродства  $A2 \rightarrow B1$  и  $A2 \rightarrow B2$  в точности одинаковы (верхняя строка). Благодаря взаимодействиям  $CS$  обеих связей становятся очень малыми и могут вообще «не выжить». Если исходные степени сродства элемента  $A2$  не в точности равны друг другу, то обычно более сильная из них выдер-

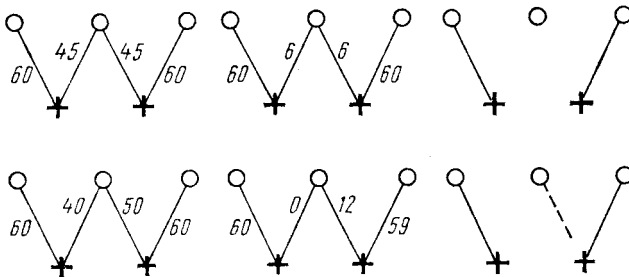


Рис. 2.18. Несимметричное движение

живает конкуренцию (нижняя строка). Из этого следует такой вывод: симметричное сопоставление возможно, но неустойчиво, а потому и маловероятно. На рис. 2.18 показаны результирующие значения  $CS$  (при малом МСИ) как для симметричного, так и для несимметричного случаев.

**Пример 4: симметричное движение.** Три элемента, расположенные в строку, предъявляются попеременно с четырьмя другими элементами (рис. 2.19).

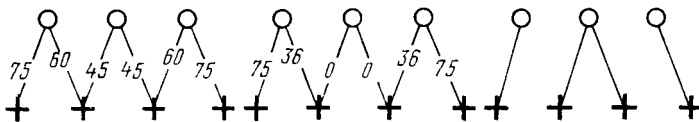


Рис. 2.19. Симметричное движение

Здесь в противоположность примеру 3 центральный элемент участвует в расщепленном движении. Для объяснения этого эксперимента привлекались глобальные принципы симметрии [Этгнев, 1974]. Однако никакие принципы, кроме локальной схемы состязаний, здесь не нужны. Значение  $CS$ , вычисленное по этой схеме, дает в точности тот вид движения, который наблюдается на самом деле.

Приведенные выше примеры подкрепляют ту точку зрения, что установление соответствия при движении происходит на основе исходных степеней сродства с помощью простых локальных взаимодействий. Схема состязаний может приемлемо объяснить наблюдаемое поведение процесса установления соответствия на исследуемых конфигурациях при малых и больших МСИ. Хотя конкретные значения, даваемые схемой, зависят от конкретных используемых функций, но качественное поведение (т. е. распределение предпочтений) определяется прежде всего основными свойствами схемы состязаний. Это значит, что функции можно менять в широких пределах, не оказывая при этом большого влияния на предсказываемое схемой установление соответствия [Ульман, 1977а; приложение 4].

## 2.4.2. Образование единиц более высокого уровня

В этом разделе будет приведен пример изображения, на котором схема состязаний дает неверный результат, если она применяется к индивидуальным элементам изображения. Эта же схема, однако, работает вполне успешно, если провести подходящее расширение множества опорных символов установления соответствия при движении.

Рассмотрим экспериментальное изображение, представленное на рис. 2.20а, б. На рис. 2.20а маленький отрезок *C* показан в

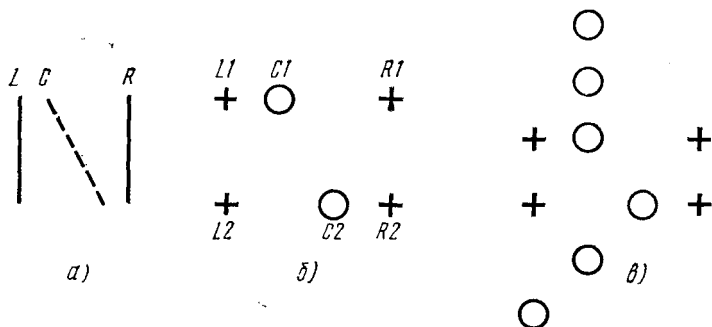


Рис. 2.20. Групповые опорные символы: пары точек (б) ведут себя так же, как отрезки (а), фоновые точки разрушают эффект группировки (в)

конкурирующем движении с отрезками *L* и *R*. На рис. 2.20,б каждый отрезок заменен парой точек — его концов. При применении к отдельным точкам на рис. 2.20,б схема состязаний, предназначенная для установления соответствия, предсказывает, что преобладать будут сопоставления  $C1 \rightarrow L1$ ,  $C2 \rightarrow R2$ . Однако при предъявлении этого экспериментального изображения такое предсказание не подтверждается. Две точки *C1* и *C2* не разделяются. Иногда они движутся налево, сопоставляясь с парой *L1*, *L2*, иногда — направо, сопоставляясь с *R1*, *R2*. Иногда эти движения воспринимаются одновременно. Способы восприятия движения и их относительная сила согласуются с изображением на рис. 2.20а, где использованы не пары точек, а отрезки.

Наша схема состязаний не в состоянии предсказать верный результат в случае применения ее к отдельным точкам. Этот пример наводит на мысль, что при описанных условиях вычислительные операции, управляющие установлением соответствия, считают каждую пару точек эквивалентом короткого отрезка. Тот факт, что в процессе установления соответствия пара пространственно близких точек обрабатывается как короткий отрезок, не должен казаться нам странным. В гл. 1 говорилось, что соответствие устанавливается между такими опорными символами, как линии, края и пятна, которые образуют строительные блоки сложных, структурированных фигур. Зрительная система обнаруживает контуры, границы и элементы текстуры с помощью различных

способов. Один из них состоит в группировании близлежащих точек или в общем случае в образовании опорного символа области, удовлетворяющего определенным условиям [Marr, 1976]. Другой возможный метод заключается в нахождении границ между областями с разной текстурой, и имеются сведения [Рамачандран и другие, 1973] о том, что движение может возникать между такими линиями раздела. При определенных условиях соответствие может установиться и между двумя размытыми пятнами или скоплениями большого числа малых элементов, и при этом не обязательно происходит сопоставление индивидуальных элементов. Исследование Юлеша и Пэйна (1968) показывает, что на стереопарах из случайных точек опорными символами при установлении движения могут также служить и границы диспаратности. Края, линии, пятна и другие элементарные опорные символы принадлежат тому уровню, который был предложен в гл. 1 для выполнения операции установления соответствия. Поэтому такой группировки следовало ожидать, но с изучением ее необходимо было повременить до разработки схемы состязаний. Не развину сему, трудно утверждать о существовании группировки, поскольку не ясно, можно ли, не прибегая к группировке, объяснить наблюдаемое в эксперименте установление соответствия. Например, большинство экспериментов гештальтпсихологии в противоположность рис. 2.20б на самом деле вообще не предполагает наличия какой бы то ни было организации, ибо они могут быть объяснены с помощью одной лишь схемы состязания (см. рис. 2.1а, б, в то время как рис. 2.1г, е могут быть результатами группировки). На рис. 2.20б движение элементов зависит не только от простого критерия пространственной близости, но и от группировки этих элементов в некоторое единое образование. Например, на конфигурации, приведенной на рис. 2.20в, лишь изредка эти две точки движутся вместе, поскольку они больше не сгруппированы в один и тот же опорный символ. Второй возможный способ разрушить эффект группировки — это уменьшить время предъявления первого кадра. Впрочем, малое время предъявления создает нежелательный эффект — движение воспринимается с большим трудом.

Опорные элементы, сформированные из других опорных элементов их группировкой, будем называть *групповыми опорными символами*. Важное свойство групповых опорных символов состоит в том, что они не замещают своих составных частей, а, напротив, добавляются к набору возможных сопоставлений, участвуя в конкуренции с соответствиями, устанавливаемыми между этими составными частями, а не исключая их из игры. Если значение  $CS$  между отдельными составными частями достаточно велико, то оно может разрушить групповой опорный символ. Это верно как для отрезков (см. эксперимент «разбитое колесо» в разд. 1.3), так и для рядов точек, и чем плотнее этот ряд, тем он становится более устойчивым по отношению к такому разрушению. Истинную природу взаимодействий между соответствием групповых

опорных символов и взаимным сопоставлением их составных частей еще предстоит уточнить. Например, пока еще не ясно, может ли большое значение *CS* между группами каким-либо образом усиливать значение *CS* между их составными частями.

Подведем итог обсуждению элементов более высокого уровня.

1. Множество опорных элементов содержит групповые опорные символы. Показано, что это положение является правдоподобным допущением, объясняющим сходство в случаях, изображенных на рис. 2.20а, б.

2. Можно создавать групповые опорные символы, группируя элементы в простые образования, такие как отрезки прямых, и возможно округлые пятна и края. В противоположность понятию «поле» из области теории гештальта здесь нет никаких признаков присутствия более высокой степени организации (и даже есть свидетельства в пользу противоположного, см. гл. 1).

3. Групповые опорные символы присоединяются к множеству опорных элементов, а не подменяют своих составных частей. Следовательно, группа может двигаться как единое целое, но может также произойти и распадение на независимые движения составляющих ее элементов.

Явление группировки поднимает вопрос, имеющий отношение к смене способа восприятия конфигурации Тернуса. Может ли переход к способу восприятия по связанному элементу являться результатом группирования точек в один отрезок? В самом деле, влияние как группирования, так и выравнивания за счет МСИ состоит в том, что усиливается способ восприятия по связанному элементу, и, таким образом, за смену способов восприятия может быть ответственно любое из этих явлений. Можно, однако, показать, что само по себе выравнивание за счет МСИ достаточно, по всей видимости, для того, чтобы вызвать смену способов восприятия. Для проверки этого к фону экспериментального изображения можно добавить дополнительные точки, чтобы разрушить группировку (ср. рис. 2.20в с рис. 2.20б). На рис. 2.21 приведен пример применения этой процедуры к конфигурации Тернуса. Квадраты обозначают добавленные фоновые элементы; они входят как в первый, так и во второй кадры (при действительном предъявлении изображения все элементы были совершенно одинаковой формы). В этой конфигурации элементы *A*, *B* и *C* обнаруживают ту же самую смену способов восприятия, что и в стандартной конфигурации Тернуса. Если проводилось сопоставление групповых опорных символов, то можно было бы ожидать, что

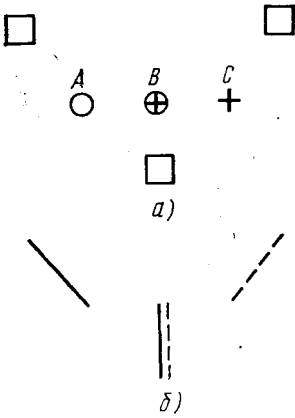


Рис. 2.21. Конфигурация Тернуса вместе с фоновыми элементами (а) и ее вид при замене пар близких точек на отрезки (б)

можно было бы ожидать, что

группировка примет форму, показанную на рис. 2.21б: близкие элементы группируются в отрезки (сплошные линии обозначают первый кадр, пунктирные — второй). При реальном исследовании движения конфигурации рис. 2.21б преимущественно устанавливается соответствие  $A \rightarrow B$  (для малого МСИ), отличающееся от сопоставления на рис. 2.21а. Из этого следует вывод, что для совершения смены способа восприятия в конфигурации Тернуса группирование не требуется. И тем не менее, остается еще возможность того, что группирование также вносит свой вклад в смену способов восприятия как в конфигурации Тернуса, так и в других конфигурациях. Мне кажется, что процессы группировки имеют особое значение и что в предстоящих исследованиях процесса установления соответствия они должны быть отнесены к главным изучаемым вопросам. Среди других проблем, требующих дальнейшей разработки, — экспериментальное определение родства и функций взаимодействия. При дальнейшем изучении различных аспектов процесса установления соответствия полезно перейти от общих вопросов к более частным. Руководствуясь этим принципом, в следующем разделе рассмотрим общий вопрос о связи между процессом установления соответствия и трехмерной интерпретацией изображений.

## 2.5. Сродство и трехмерная интерпретация

### 2.5.1. Сродство и расстояние

Расстояние между элементами является параметром, играющим важную роль при определении сродства между ними. В данном разделе ставится вопрос о том, какой вид пространственного расстояния оказывает влияние на сродство: только неинтерпретированное, двумерное расстояние или же интерпретированное трехмерное.

В процессе зрительного анализа воспринимаемым фигурам, как правило, придается трехмерная интерпретация, влияющая на оценку пространственных отношений составных частей фигуры. Зачастую результирующий перцепт определяется как двумерными, так и проинтерпретированными трехмерными связями (см. пример на рис. 2.22). Создается впечатление, что отрезок  $R$  лежит выше  $S$  и ниже  $L$ . В трехмерном пространстве  $S$  и  $R$  находятся на одном уровне, если для представления вертикали используется направление  $S$  и  $R$ . В двумерной плоскости изображения  $R$  и  $L$  в точности симметричны относительно  $S$  (и  $D$ ). Таким образом, воспринимаемая разница в высоте не согласуется ни с трех-, ни с двумерными отношениями, но находится под влиянием тех и других.

Воспринимаемое расстояние (т. е. расстояние в плоскости изображения, оцененное наблюдателем) также находится под влиянием трехмерной интерпретации, так что воспринимаемые расстояния, например, между  $D$  и  $L$  по сравнению с расстоянием

между  $D$  и  $R$  будут чем-то средним между чисто двумерным и интерпретированным трехмерным расстояниями. Таким образом, можно различить три типа расстояний: 1) двумерное неинтерпретированное; 2) трехмерное, полученное в результате пространственной интерпретации сцены; 3) воспринимаемое, обусловленное и двумерным, и трехмерным расстояниями (ср. с работой [Эттив и Фрост, 1969] и с «феноменальным разнесением» в [Эттив и Блок, 1973]). В двух последних случаях будем говорить, что установление соответствия зависит от *проинтерпретированного*, тогда как в первом случае оно зависит от *неинтерпретированного* расстояния. Рассматриваемая проблема состоит в выяснении того, какое из этих трех расстояний подходит для процесса установления соответствия.

Для детального изучения проблемы имеется несколько причин. Главная из них состоит в том, что эта проблема является важным и общим аспектом проблемы установления соответствия, который надо прояснить, прежде чем вдаваться в детали. В процессе сопоставления велико влияние расстояния, значит, необходимо уточнить термин «расстояние». Эта проблема имеет также прямое отношение к тому *уровню*, на котором выполняется установление соответствия.

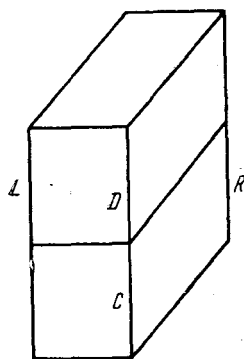


Рис. 2.22. Отрезки  $L$  и  $R$  в точности симметричны относительно  $C$  и  $D$ ; по гипотезе неинтерпретированного расстояния пары  $C-R$  и  $C-L$  должны иметь одинаковые  $CS$

Если формирование пространственной модели сцены происходит раньше процесса установления соответствия, то предпочтение, отдаваемое при сопоставлении с ближайшими соседями, могло бы быть основано на трехмерных расстояниях в этой модели. Однако на основании предыдущих разделов следует ожидать, что соответствие устанавливается независимо от трехмерной интерпретации, и в силу этого процессом сопоставления должно управлять неинтерпретированное расстояние. По-видимому, это противоречит экспериментальным данным, приводимым в литературе по психологии. Исследования Корбэна (1942), Эттивна и Блока (1973) истолковывались как подтверждение той точки зрения, что проинтерпретированное расстояние лежит в основе процесса установления соответствия или, по крайней мере, оказывает на него сильное влияние. Однако все предшествующие исследования основывались на парадигме оценок гладкости, применимость кото-

рой к проблеме установления соответствия, как мы уже отмечали, сомнительна. Следовательно, второй целью данного раздела является критическое рассмотрение связей между гладкостью движения и степенью соответствия.

Наши выводы основаны на результатах экспериментов, приведенных в работе [Ульман, 1978а]. При определении того, какой вид расстояния управляет процессом установления соответствия,

использовались прямые сравнения значений степени соответствия. Были придуманы разнообразные фигуры, содержащие отрезок прямой, чередующийся с двумя другими отрезками, из которых один был расположен слева, а другой — справа от первого. На некоторых экспериментальных изображениях эти линии были включены в конфигурации, имевшие явное трехмерное истолкование. Затем для проверки различных гипотез о расстояниях анализировалось воспринимаемое движение. В одном из примеров отрезок  $D$  на рис. 2.22 испытывался в условиях конкурирующего движения с отрезками  $L$  и  $R$ . Благодаря трехмерной интерпретации этой фигуры расстояния между  $D$  и  $L$ ,  $D$  и  $R$  выглядели заметно отличающимися друг от друга. Если трехмерное внутреннее представление фигуры действительно формируется до проведения операции сопоставления, то преобладать должно движение между  $D$  и  $L$ , поскольку в модели расстояние между ними будет меньше расстояния между  $D$  и  $R$ . Следовательно, предсказание, основывающееся на гипотезах об интерпретированном расстоянии, состоит в том, что если  $D$  предьявляется в условиях кажущегося движения одновременно с  $L$  и  $R$ , то движение в сторону  $L$  будет оценено наблюдателем как более сильное, причем на величину, которая оценена в другом, связанном с этим эксперименте. Поскольку в плоскости изображения расстояния между  $D$  и  $L$ ,  $D$  и  $R$  равны между собой, то гипотеза о неинтерпретированном расстоянии предсказывает, что два конкурирующих движения будут равновероятны. Когда отрезки были испытаны на движение в действительности, то оказалось, что оба конкурирующих движения возникают с одинаковой вероятностью. Результаты этого и ряда других экспериментов приводят к заключению, что именно неинтерпретированное расстояние управляет процессом установления соответствия.

Эти результаты были получены с помощью методики конкурирующих движений, которая сравнивает между собой значения  $CS$  непосредственно. Как отмечалось выше, в экспериментах, похожих на этот, но с применением критерия гладкости движения, были получены иные результаты. Поэтому интересно более подробно изучить эти различия.

### 2.5.2. Степень соответствия и гладкость движения

Межстимульный интервал, требующийся для возникновения оптимального\* кажущегося движения, зависит от пространственного расстояния между стимулами, что первоначально было сформулировано в виде третьего закона Корте [Корте, 1915]. Является ли расстояние, подходящее для этого случая, двумерным или перцептивным? Корбэн (1942) и Эттнив с Блоком (1973) обратились к экспериментальному решению этого вопроса. Они выяснили, что гладкость (или оптимальность) движения определяется преимущественно (а может быть, и только) перцептивным рас-

\* То есть наиболее ясно воспринимаемого движения. — Прим. ред.



стоянием. Колерс (1972, гл. 4 и 5), как и более ранние исследователи [например, Браун и Вот (1937)], изучал относящиеся к  $CS$  проблемы, используя гладкость движения как критерий для  $CS$ . Но предположение о том, что гладкость движения является мерой степени соответствия, не согласуется с указанными выше результатами. Несовместными являются следующие три предположения: 1) гладкость движения зависит от перцептивного расстояния; 2)  $CS$  зависит от двумерного расстояния; 3) гладкость движения является мерой  $CS$ .

Первые два предположения были подвергнуты экспериментальной проверке, а третье — нет. Поэтому желательно проверить третье предположение непосредственно. Проверка его может быть выполнена следующим образом. Сперва попеременно показываются линии  $C$  и  $L$  ( $L$  параллельна  $C$  и лежит от нее слева) и измеряются временные интервалы, необходимые для возникновения оптимального движения. Тем самым определяется временной интервал  $(t_1, t_2)$ , обладающий таким свойством: если МСИ удовлетворяет требованию  $t_1 < \text{МСИ} < t_2$ , то  $C$  и  $L$  создают ощущение гладкого движения. Затем та же процедура повторяется для второй пары линий  $C$  и  $R$ , причем  $R$  лежит справа от  $C$  и расстояние от  $C$  до  $R$  больше расстояния от  $C$  до  $L$  (само  $L$  в это экспериментальное изображение не входит). Тем самым для пары  $C$  и  $R$  получается новое множество интервалов  $(T_1, T_2)$ . Наконец,  $C$  попеременно показывается с  $L$  и  $R$  вместе. При этом межстимульный интервал удовлетворяет условиям:  $T_1 < \text{МСИ} < T_2$ ;  $t_2 < \text{МСИ}$ . Такой МСИ попадает в область гладкости при движении пары  $C$  и  $R$ , но не  $C$  и  $L$ . Какой перцепт должен при этом возникнуть? Если руководствоваться сделанным выше предположением 3), то преимущественным должно бы быть движение пары  $C$  и  $R$ .

Такой эксперимент был поставлен, и его результаты описаны в работе [Ульман, 1978а]. Интервалы гладкого движения определялись для пар изолированных вертикальных прямых (т. е. без всякого фонового окружения). Расстояние между  $C$  и  $L$  составляло  $0,7^\circ$  зрительного угла, а  $C$  и  $R$  отстояли друг от друга на  $4,5^\circ$ . Среднее значение  $t_2$  (верхний предел диапазона гладкости для более близких прямых) составляло 115,4 мс (со стандартным отклонением  $CO=26,2$ ), а для  $T_2$  среднее значение было 172,7 мс ( $CO=34$ ). Впрочем, эти средние значения не были использованы в эксперименте. Вместо этого для каждого испытуемого индивидуально определялись диапазоны возникновения оптимального движения, и затем вычислялся МСИ, удовлетворяющий вышеприведенным требованиям. Из 12 участвовавших в эксперименте наблюдателей десять отметили движение только между  $C$  и  $L$ , один отдал большее предпочтение движению влево и один — меньшее предпочтение движению влево.

Вывод: отказом от третьего предположения снимается возникшее в этом разделе противоречие. Гладкость движения *не может* служить показателем степени соответствия.

### 2.5.3. Общие выводы

Из приведенных экспериментов следуют два основных результата: первый относится к сути проблемы установления соответствия, второй — к связи между гладкостью движения и степенью соответствия. Тот факт, что гладкость не является истинной мерой  $CS$ , предостерегает нас от того, чтобы сосредоточиваться на изучении кажущегося движения путем оценок «оптимальности движения». До тех пор, по крайней мере, пока речь идет о проблеме установления соответствия, рассмотрения, основанные на оптимальности движения, могут вводить нас в заблуждение, и, таким образом, большое число исследований, основанных на гладкости движения, нельзя признать приложимым к изучаемой проблеме.

Применяя методику конкурирующих движений, в которой непосредственно сравниваются между собой значения  $CS$ , мы пришли к выводу, что эти значения определяются двумерным расстоянием. В рамках полученного результата стоит обратить внимание на две отличительные черты. Во-первых, двумерное расстояние вовсе не обязательно должно быть сопоставимо с расстоянием на сетчатке глаза. То, что эти два расстояния действительно могут быть различны, было показано Рокком и Эбенхолтцем (1962). Во-вторых, из этих экспериментов была исключена глубина, получающаяся из данных о диспаратности, поскольку диспаратность может быть определена на нижнем уровне независимо от интерпретации сцены, как это было продемонстрировано Олшем (1971). Возможное влияние стереоскопической информации на  $CS$  еще предстоит выяснить в ходе дальнейших исследований.

### 2.5.4. Семантическая интерпретация

Влияние семантической интерпретации на соответствие, устанавливающееся между движущимися объектами, было изучено в нескольких работах. Джоунс и Брунер (1954), к примеру, сравнивали движение, воспринимаемое между схематичными фигурками людей, составленными из отрезков, с движением между бессмысленными фигурами. Однако поскольку это исследование было основано на сравнениях гладкости, то его нельзя считать относящимся к проблеме установления соответствия.

Работа, изучающая влияния семантической интерпретации и использующая методику конкурирующих движений, а не оценки гладкости, была выполнена Тохом и Иттелсоном (1956). В этой работе было отмечено влияние семантической интерпретации на процесс сопоставления. Эксперимент проводился с двумя изображениями. На одном экспериментальном изображении рисунок направленной вниз бомбы испытывался в условиях конкурирующего движения с рисунками двух других бомб, расположенных

выше и ниже первой. На втором изображении рисунок направленного вверх самолета испытывался в условиях конкурирующего движения с самолетами, расположенными над и под ним. Утверждалось, что для бомбы несколько более предпочтительным было движение вниз, а для самолета — вверх, как этого и можно было бы ожидать по смыслу сцены.

Есть, однако, причины не считать указанный результат относящимся к *CS*. Главная причина — недостаточный контроль за движениями глаз. Можно ожидать, что получатся аналогичные результаты, если глаза испытуемого будут перемещаться соответственно предполагаемому направлению движения (ср. с «поездом из кубиков» в разд. 1.3). Эксперименты проводились с изображениями, составленными из стрелок, направленных в разные стороны, и пока испытуемый фиксировал свой взгляд на центре изображения, не было обнаружено никакого влияния семантической интерпретации. К проблеме семантической интерпретации и установления соответствия при движении также имеют отношения некоторые исследования Нэвона (1976), в которых была использована методика конкурирующих движений. Хотя Нэвон интересовался сходством фигур, но некоторые из его экспериментальных изображений осуществляют также и сравнение между семантически тождественными и семантически различными элементами (например, английскими буквами) в условиях конкурирующих движений. Сообщаемые им результаты не обнаруживают никакого влияния семантического сходства на образование пар при сопоставлении и, таким образом, подкрепляют ту точку зрения, что установление соответствия при движении не зависит от семантической интерпретации. В заключение можно сказать, что единственным доказуемым влиянием, оказываемым семантической интерпретацией на процесс установления соответствия, является косвенное влияние, осуществляемое с помощью движений глаз. Если глаза остаются неподвижными, то семантическая интерпретация, как и трехмерная, не оказывает никакого влияния на степень соответствия.

Имеющиеся данные подкрепляют тот взгляд, что процесс установления соответствия является операцией нижнего уровня, которая предшествует интерпретации сцены или не зависит от нее. Такая операция нижнего уровня в состоянии объяснить воспринимаемую глазом целостность объектов, находящихся в движении или изменении. Зрительная система человека способна правильно отождествить между собой различные проекции одного движущегося объекта как представляющие одно и то же физическое образование. Изменяющиеся объекты также могут сохранять идентичность: в мультфильме лягушка может прямо на наших глазах превратиться в царевну, не теряя идентичности самой себе. Можно предположить, что эта идентичность сохраняется потому, что процесс установления соответствия, работая на нижнем уровне, объявляет эти изображения соответствующими друг другу, т. е. различными представлениями одной и той же физической

уности независимо от их последующей трехмерной и семантической интерпретаций.

Нижеуровневый автономный процесс установления соответствия может быть также использован другими процессами, возникающими в ходе восприятия. Эта возможность обсуждается в следующем разделе.

## **2.6. Возможность применения процесса установления соответствия к возникновению представлений об объектах**

В этом разделе ставится вопрос о возможности того, что процесс установления соответствия может играть важную роль на ранних этапах формирования представлений об объектах. Наша цель заключается не в том, чтобы сделать хорошо обоснованное предположение, поскольку для этого у нас еще не хватает данных, а скорее в том, чтобы указать на тот факт, что наличие автономного процесса установления соответствия может повлечь за собой выводы, относящиеся к восприятию объектов и образованию представлений о них (их концептуализации). Укажем на возможную схему, которая в значительной степени оказалась вне поля зрения исследователей, вероятно, потому, что процесс установления соответствия обычно не рассматривался как процесс элементарный и автономный. Хоть и умозрительно, но в этом разделе делаются конкретные выводы, которые, вероятно, могут быть проверены экспериментально.

Формирование представлений об объектах является одним из наиболее фундаментальных аспектов развития познавательных способностей в младенческом возрасте [Бауэр, 1966, 1971, 1974; Гибсон, 1969; Дэй и Маккензи, 1973; Козн и Салапатека, 1975]. Это процесс, благодаря которому ребенок, совершенно не знакомый с материальными объектами, становится ребенком, явно обладающим знаниями об объектах и их свойствах.

Для понимания этого процесса надо разрешить две главных проблемы: 1) как он развивается во времени, а именно, что действительно знают дети о материальных объектах в разном возрасте; 2) как ребенок получает это знание на основании своего сенсорного контакта со средой.

Формирование представления об объекте многогранно, и одной из его сторон является способность связывать между собой разные изображения, принадлежащие одному и тому же объекту. При попытке объяснить эту способность возникают две взаимоисключающие точки зрения: связывание изображений может быть основано на какой-то зачаточной системе представлений об объектах либо соотношение может быть обратным и возникновение представлений об объектах является следствием способности к связыванию изображений. Согласно первой точке зрения сначала ребенок может как-то понять тот факт, что определенное изображение  $V_1$ , которое предъявляется ему, представляет некоторый объект  $O_1$ , а другое изображение  $V_2$  — объект  $O_2$ . Изображения

$V1$  и  $V2$  связываются между собой только тогда, когда ребенок на основе, быть может, осязания приходит к осознанию того факта, что  $O1$  и  $O2$  являются одним и тем же объектом. При таком подходе связывание изображений основано на существовании некоторого зачаточного представления об объектах, и потому может возникнуть порочный круг, если объяснять возникновение представлений об объектах через представления об объектах. Наоборот, может быть так, что  $V1$  и  $V2$  спонтанно связываются в одну группу, которая в свою очередь служит основой для зарождающегося представления об объекте. При таком подходе к проблеме формирования представлений требуется, чтобы ребенок производил две основные операции: спонтанное связывание в одну группу различных изображений одного объекта и приписывание свойства группе в целом. Объясним это положение на примере.

Предположим, что если ребенок видит различные изображения данного объекта (скажем, молочной бутылочки), то он может сгруппировать их некоторым образом, который можно графически представить как создание некоторого образования более высокого уровня в его системе представления мира (рис. 2.23а). Способность к приписыванию свойств всей группе в целом можно обеспечить, приписывая их новому «узлу» более высокого порядка, даже если эти свойства получены в результате обучения на одном конкретном изображении или наборе изображений, как это показано на рис. 2.23б.

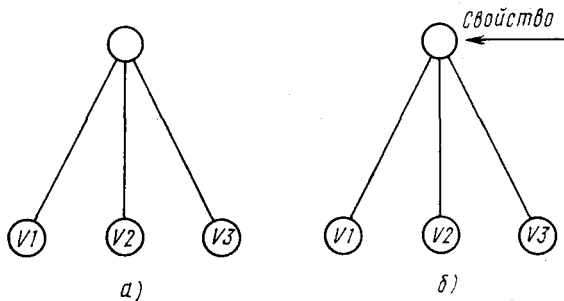


Рис. 2.23. Различные изображения связываются в узел более высокого порядка (а); свойства приписываются узлу более высокого порядка и тем самым сразу всем этим изображениям (б)

Чтобы оценить следствия, вытекающие из наличия этих двух операций, рассмотрим такой сценарий. Ребенку дают поиграть с пустой бутылочкой, так что он видит ее в различных проекциях. Когда он плачет от голода, ему вновь дают бутылочку, но уже с едой и таким способом, что по пути ко рту ребенка бутылочка видна только в некоторой проекции из всего множества возможных проекций. По прошествии некоторого времени ребенок, вероятно, обучится связывать ожидание пищи с этим изображением бутылочки. Если он обладает двумя перечисленными выше опе-

ациями, то ожидание пищи будет связываться не с какими-то изображениями бутылочки, а с узлом более высокого порядка и, следовательно, сразу со всеми изображениями бутылочки. Бутылочка в одном ракурсе будет вызывать у голодного ребенка ту же реакцию, что и в любом другом, хотя в прошлом пища ассоциировалась только с некоторым подмножеством изображений бутылочки. Можно сказать, что такой ребенок имеет первичное понятие о молочной бутылочке. Его поведение можно описать как осознание до некоторой степени того факта, что свойство «обеспечивать пищей» относится к объекту «бутылочка», а не к частным ее изображениям.

Следовательно, предположение о том, что связывание изображений друг с другом происходит до формирования первоначального представления об объекте, а не после него, несколько более приемлемо. Если бы способность различать объекты и их изображения и была необходимой предпосылкой для связывания друг с другом разных проекций одного и того же объекта, то все равно стало бы непонятным, в чем состоит эта способность и как она проявляется. Если же процесс группирования изображений не зависит от какого-либо предварительного знания об объектах, то это дает нам возможность разорвать «порочный круг», возникающий при обосновании появления представлений об объектах существованием каких-то первоначальных представлений о них. При этом также нет необходимости предполагать, чтобы грудной младенец сколь-нибудь глубоко понимал, чем на самом деле является объект внешнего мира. Два основных процесса (связывание изображений и приписывание свойств) по сути своей могут оставлять первый этап в понимании того, чем является некоторый объект.

Обрисованная выше общая схема формирования представлений об объектах требует, чтобы различные изображения данного объекта, быть может, совсем несхожие, связывались друг с другом, но не с изображениями других объектов. Но как можно выделить различные ракурсы некоторого объекта, не зная, что это за объект? В прошлом были предложены различные механизмы отбора и связывания между собой изображений, хотя авторы и не всегда явно формулировали, что они предназначены для решения именно этой задачи. Наиболее существенными здесь являются ассоциации или многомодалное восприятие по Пиаже и введенное Бауэром понятие постоянного местоположения.

Ассоцианисты — от Локка до Дж. С. Милля — полагали, что совокупности чувственных впечатлений группируются в кластеры на основе таких критериев, как одновременность, пространственная близость и сходство. С точки зрения Пиаже (1954) выработка представления об объекте зависит от обучения, вовлекающего восприятия различной модальности, она, в частности, зависит от комбинации зрительного и осязательного исследований объекта. Так, ощущение удерживания некоторого объекта может дать информацию о том, что он остается одним и тем же, хотя при мани-

пуляции внешний вид его изменяется. Бауэр (1971) выдвинул соображение, что тождественность объекта самому себе устанавливается в первую очередь на основе его местоположения. Если объект изменяется внешне, но остается на прежнем месте, то различные его ракурсы будут восприниматься, по крайней мере в некотором смысле, как одни и те же.

Не будем отвлекаться на рассмотрение этих гипотез, а рассмотрим новый вариант, который может быть особенно уместным в нашей задаче. Два изображения могут быть связаны между собой, если они *находятся в отношении соответствия*, т. е. если одно из них преобразуется в другое с помощью движения.

В предыдущих разделах приводились доводы в пользу того, что установление соответствия представляет собой автономный процесс нижнего уровня, который функционирует независимо от предварительного знакомства с объектами и их распознавания. Повседневный опыт иллюстрирует успешную работу процесса установления соответствия. Его адекватность проявляется в том, что объекты, в том числе и движущиеся, не теряют перцептивной идентичности самим себе и не сливаются с другими объектами. Поэтому можно ожидать, что всякий раз, когда два ракурса сопоставляются друг с другом с помощью процесса установления соответствия на основе значений их сродства и определенных пространственно-временных отношений, оба ракурса действительно представляют один и тот же объект. Следовательно, их спокойно можно связать между собой для образования требуемой основы развития представления об объекте.

Предложенная выше гипотеза может быть проверена экспериментально. К примеру, ребенку предъявляют изображение *V1* некоторого объекта, которое путем гладкого движения преобразуется в другое изображение *V2*. Затем ребенка приучают связывать с *V1* определенное свойство и исследуют перенос обучения на *V2*. Этот перенос, основанный на движении объекта, можно сравнить с другим эффектом, возникающим, когда *V1* и *V2* следуют друг за другом в быстрой временной последовательности, но не воспринимаются наблюдателем в гладком движении. Предсказание, основанное на движении, состоит в том, что указанный перенос свойства с одного изображения объекта на другое произойдет значительно легче в первом случае, когда изменения вида объекта обусловлены его движением.

В данном разделе выдвинута следующая точка зрения на отношение между представлением об объекте и анализом движущихся объектов. Вместо ситуации, в которой исходными берутся представление о неподвижном объекте и его различные ракурсы, а затем возникшая структура используется для анализа движущихся объектов, может существовать прямо противоположная ситуация: вначале разные изображения объекта отождествляются друг с другом на основе установления соответствия при движении, и уже тем самым выявляется структура представления как основа понимания того, чем же является объект.

## ТЕОРИЯ УСТАНОВЛЕНИЯ СООТВЕТСТВИЯ ПРИ ДВИЖЕНИИ, ОСНОВАННАЯ НА МИНИМАЛЬНОМ ОТОБРАЖЕНИИ

### 3.1. Введение

В этой главе будет рассмотрен вопрос о том, каким образом может быть успешно установлено соответствие при движении, и подойдем к проблеме установления соответствия с вычислительной точки зрения. Наша цель состоит в том, чтобы разработать подходящую схему установления соответствия и объяснить имеющиеся у нас данные об установлении соответствия зрительной системой человека.

Множество подлежащих рассмотрению возможных стратегий установления соответствия в значительной степени определяется тем, на каком уровне осуществляется процесс сопоставления. Другими словами, стратегия зависит от тех элементов, которые должны быть сопоставлены, например от того, являются ли они образованиями высокого уровня, такими как воспринимаемые объекты [ср. с Уоррен, 1977; Ульман, 1977b], или единицами низкого уровня, такими как точки, пятна, фрагменты краев, отрезки и некоторые группы этих элементов. В гл. 1 и 2 приведены выводы и экспериментальные данные, поддерживающие вторую точку зрения. Таким образом, в случае процесса установления соответствия входное изображение в любой фиксированный момент можно рассматривать просто как набор элементарных опорных символов, а теория сопоставления этих символов друг другу должна создать достаточную основу для общей теории установления соответствия при движении.

Начнем с рассмотрения упрощенного варианта проблемы установления соответствия, в котором два кадра предъявляются последовательно, приводя к возникновению кажущегося движения. Кроме того, будем считать, что каждый кадр состоит только из изолированных точек одинаковой яркости. Впоследствии этот анализ будет распространен на другие типы элементов и непрерывное движение.

В поисках подходящей процедуры установления соответствия будем использовать так называемые указания сверху и запреты снизу. Указания сверху — это свойства физического движения, которые можно использовать для решения задачи установления соответствия. Запреты снизу — это ограничения, налагаемые общими вычислительными соображениями и в особенности требованиями осуществимости этого вычисления в биологической системе.

### 3.2. Оптимальная (независимая) стратегия установления соответствия

При наличии двух кадров приходится сталкиваться с проблемой: как установить соответствие между их элементами? Если



считать, что каждый кадр содержит  $n$  элементов, то различных взаимно-однозначных соответствий между кадрами будет  $n!$ . Следовательно, возникает проблема неоднозначности, общая для разных сторон зрительного анализа, а именно проблема, состоящая в том, что входной зрительный сигнал допускает более чем одну интерпретацию. В свете этой неоднозначности ни один из методов не застрахован от неверной интерпретации. Если, однако, какие-то свойства окружающей наблюдателя среды делают некоторые интерпретации правдоподобнее других, то становится возможным отобрать наиболее правдоподобное решение, максимизируя, таким образом, вероятность того, что входная информация проинтерпретирована правильно. Поэтому будем искать такую схему установления соответствия, которая бы максимизировала вероятность получения правильной интерпретации<sup>3.1</sup>.

Выбор наиболее правдоподобного соответствия требует использования информации о правдоподобии различных вариантов сопоставления. Эта дополнительная информация может быть общей и частной. В случае частной информации используются сведения, применимые к особой ситуации, например предположение о том, что черное пятно на столе в конторе представляет собой телефон. Примерами сведений общего характера являются ограничение, налагаемое неизменностью формы тела на процесс интерпретации структуры по данным о движении (вводится в гл. 4), или два ограничения, регулирующие процесс сопоставления при стереоскопическом восприятии [Марр и Поджио, 1976]. Если при движении соответствие устанавливается на нижнем уровне, то должна применяться информация общего характера. В следующем разделе для регулирования процесса сопоставления будут применены общие свойства движущихся элементов.

**Гипотеза независимости.** Для отбора наиболее оправданного соответствия требуется метод сравнения правдоподобия различных вариантов сопоставления. Для того чтобы определить правдоподобие некоторого варианта сопоставления, необходимы сведения о предполагаемых взаимоподчинениях, которые имеются при движении отдельных элементов. Если, к примеру,  $X$  и  $Y$  — соседние элементы и  $X$  движется направо, то будет ли для  $Y$  движение направо более вероятным, чем налево? Поскольку наша главная цель состоит в исследовании зрительного восприятия человека, то необходимо, чтобы основные предположения были согласованы с тем, как зрительная система человека выполняет процесс установления соответствия. При изучении процесса установления соответствия для зрительной системы человека с использованием простых экспериментальных изображений, содержащих малое число элементов, подобные смещения вероятности не наблюдаются. На рис. 3.1 демонстрируется отсутствие такого смещения. Точки  $X1$  и  $X2$  предъявляются в условиях кажущегося движения с точками  $Y1$ ,  $Y2$  и  $Y3$ . (На всех рисунках пустые кружки будут обозначать первое предъявление, крестики — второе; крестик внутри пустого кружка означает, что данный эле-

мент входит как в первый, так и во второй кадры.) Если предъявляются только  $X1$ ,  $Y1$  и  $Y2$ , то  $X1$  с одинаковой вероятностью движется вправо и влево. Как правило, он расщепляется и движется сразу в двух направлениях. Если вместе с ними предъявляются также  $X2$  с  $Y3$ , то видно, как  $X2$  движется вправо и сопоставляется с  $Y3$ . Будет ли это движение повышать вероятность того, что мы увидим  $X1$  движущимся направо и сопоставляющим с  $Y2$ ? Факт подобного предпочтения не наблюдается при условии поддержания фиксации взгляда на центре изображения.

В гл. 2 было показано, что для объяснения свойств процесса сопоставления необходимо привлечение конкурирующих взаимодействий, однако «положительные» взаимодействия между сопоставлениями сходных направлений не требовались. Обобщим эти наблюдения и примем гипотезу, по которой элементы движутся независимо друг от друга<sup>3,2</sup>. С учетом этой гипотезы независимости разработаем оптимальную стратегию становления соответствия. Впоследствии будет показано, что получившийся метод остается оптимальным и в условиях, когда гипотеза независимости может нарушаться, что поэтому введение взаимозависимости направлений было бы излишним.

**Максимально правдоподобное соответствие.** Предположим, что в пространстве независимо друг от друга, с разными скоростями и в разных направлениях движутся  $n$  элементов. Делается «мгновенная снимка» этих движущихся элементов, и надо установить соответствие между входными (на первом изображении) и выходными (на втором) элементами. Обозначим через  $p(v)$  распределение вероятностей скоростей элементов в плоскости изображения. Это значит, что если случайным образом выбрать движущийся элемент, то вероятность того, что его скорость в плоскости изображения лежит между значениями  $a$  и  $b$ , составляет

$$\int_a^b p(v) dv. \quad (3.1)$$

Если принять гипотезу независимости, то вероятность реализации набора из  $n$  элементов, в котором  $i$ -й элемент ( $1 \leq i \leq n$ ) за время  $t$  проходит расстояние  $d_i$ , дается произведением

$$\prod_i p(v_i), \quad (3.2)$$

где  $v_i = d_i/t$ . Поэтому наиболее правдоподобный вариант сопоставления ищется путем максимизации выражения (3.2) на множестве всех допустимых вариантов сопоставления двух кадров (в данном случае это взаимно-однозначные отображения). Для

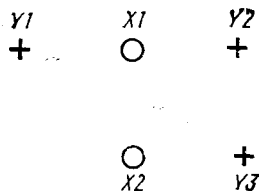


Рис. 3.1. Нехватка «положительных взаимодействий» между вариантами сопоставления

последующего нам удобно будет преобразовать произведение (3.2) в сумму. Поскольку логарифмическая функция монотонна, а  $p(v_i)$  положительны, то наиболее правдоподобный вариант сопоставления эквивалентным образом можно найти, решая задачу

$$\min \sum q(v_i), \quad (3.3)$$

где минимум берется по множеству всех допустимых вариантов сопоставления, а  $q(v) = -\log p(v)$ . Если  $0 \leq p(v) \leq 1$ , то  $q(v)$  будет неотрицательной функцией.

**Примечание:** до конца этой главы индекс суммирования  $i$  будет опускаться всюду, где это не вызовет неоднозначности.

Если  $q(v)$  рассматривать как функцию стоимости, то оптимальное отображение минимизирует общую стоимость на множестве всех допустимых вариантов сопоставления.

**Отображения, не являющиеся взаимно-однозначными.** Если число входных элементов не равно числу выходных, то отображение между ними не может быть взаимно-однозначным. Простейший пример такой ситуации дает конфигурация расщепляющегося движения (см. рис. 2.3), в которой  $X$  предьявляется в условиях кажущегося движения с  $Y1$  и  $Y2$ . В этом случае условие взаимно-однозначности должно нарушаться, что может происходить одним из двух способов: либо  $X$  отображается только в один элемент, оставляя другой без «партнера», либо расщепляется и отображается сразу в  $Y1$  и  $Y2$ . (Если  $Y1$  и  $Y2$  предьявляются раньше  $X$ , то они оба могут отобразиться в  $X$  — такую ситуацию мы назвали *слиянием*.) Возможности расщепления или слияния оказываются, как правило, предпочтительнее для восприятия, чем такая ситуация, когда вообще не происходит никакого отображения (если только одно из расстояний не будет много больше другого). Поэтому будем требовать, чтобы допустимые варианты сопоставления являлись бы *покрытиями*. Покрытие определяется как сопоставление, при котором каждый входной элемент переходит хотя бы в один выходной и в каждый выходной переходит хотя бы один входной. Как для взаимно-неоднозначных случаев должен быть определен оптимальный вариант сопоставления? Гипотеза независимости в том виде, в каком она была сформулирована выше, не может быть непосредственно применена к тем ситуациям, в которых происходит расщепление или слияние элементов. Для простоты расширим гипотезу независимости так, чтобы она относилась и к покрытиям. Некоторые дополнительные видоизменения оптимального сопоставления будут введены после приведения метода вычисления оптимальных сопоставлений. Пока что оптимальное сопоставление, как и прежде, будет определяться как сопоставление, минимизирующее стоимость  $\sum q(v_i)$  по всем допустимым вариантам сопоставления. Единственное изменение состоит в том, что множество допустимых вариантов сопоставления расширяется и включает в себя все покрытия. Определенное таким образом оптимальное сопоставление в теории графов называется минимальным взвешенным покрытием двудольного графа.

В двудольном графе множество вершин  $V$  удовлетворяют условию  $V = V_1 \cup V_2$ , где  $V_1 \cap V_2 = \emptyset$ , а также ребро графа соединяет вершину из  $V_1$  с вершиной из  $V_2$ .) Для краткости будем использовать термин *минимальное отображение* для обозначения сопоставления, минимизирующего  $\sum q_i$  на множестве покрытий графа возможных пар элементов.

Теперь обратимся к изучению возможности минимального отображения и к сравнению его с процессом установления соответствия в зрительной системе человека. В гл. 2 показано, что окончательное соответствие получается из исходных значений средства с помощью локальных взаимодействий. В следующем разделе необходимость такой схемы будет обоснована теоретически. Затем займемся такой существенной проблемой: совместимо ли минимальное отображение со схемой, основанной на средстве и локальных взаимодействиях? В разд. 4 и 5 на этот вопрос дается положительный ответ, выясняется локальная природа задачи установления соответствия и предлагается метод вычисления минимального отображения по исходным степеням средства с помощью простых локальных взаимодействий.

### 3.3. Осуществимость вычислений

Искомое минимальное отображение может быть вычислено явным образом, например, с помощью перебора: сумму  $\sum q(v_i)$  можно вычислить для всех допустимых вариантов сопоставления, а затем выбрать ее минимальное значение. Однако такой алгоритм был бы неприемлем вследствие своей неэффективности. Существует ли разумный метод вычисления минимального отображения? Осуществимость вычислений в значительной степени зависит от свойств процессора, который его производит, и потому этот вопрос нельзя решить, не сделав каких-то предположений относительно способа вычисления. Поскольку вычисление соответствия выполняется и зрительной системой человека, то пусть наша концепция вычислительной осуществимости включает в себя требование биологической осуществимости. После этого можно задать вопрос: «Есть ли у нас основания ожидать, что в зрительной системе человека происходит вычисление минимального отображения?». Не ограничивая себя рамками какой-либо конкретной модели, сделаем три общих предположения о том, как в зрительной системе человека реализуется процесс установления соответствия. Затем выясним, осуществимо ли при этих предположениях вычисление минимального отображения. Эти предположения таковы: параллелизм, простота и локальность.

**Параллелизм.** Поскольку процесс установления соответствия работает с элементами нижнего уровня, а на данном конкретном изображении их может быть очень много, то предполагаем, что связывание в пары соответствующих друг другу элементов происходит в значительной мере параллельно.

**Локальность.** Если процессоров, производящих вычисление со-ответствия, много, то невозможно соединить каждый из них со всеми остальными. Поэтому предположим, что между процессо-рами имеются только локальные связи, например, каждый про-цессор связан только с  $r$  ближайшими к нему процессорами<sup>3,3</sup>. Число  $r$  будем называть *радиусом вычисления*.

**Простота.** Если процессоров много, то представляется обос-нованным предположение о том, что каждый процессор в отдель-ности является довольно простым вычислительным устройством. Здесь мы не будем пытаться точно определить термин *простота*, однако заметим, что эта простота относится к отдельным процес-сорам, но не к процессу в целом.

Объединим эти предположения в понятие *простой сети*, обо-значающее локально-связную сеть, состоящую из простых про-цессоров. В идеальном случае все процессоры идентичны и каж-дый из них связан с  $r$  своими соседями. При вычислении соответ-ствия процессоры «назначаются» элементам изображения, и их задача состоит в нахождении (параллельно) вариантов сопостав-ления для данных элементов.

Некоторые перечисленные теоретические соображения позво-ляют считать, что установление соответствия при движении осу-ществляется простой сетью. В гл. 2 приведены свидетельства в пользу той точки зрения, что процесс установления соответствия, используемый зрительной системой человека, действительно явля-ется простым и локальным.

Такие общие проблемы, как вычислимость, эффективность и локальность, в подобных простых сетях все еще недостаточно изучены. Вместо того чтобы прямо обращаться к этим пробле-мам, ограничим наше обсуждение анализом их связи с процессом установления соответствия. Поскольку минимальное отображение было предложено в качестве оптимальной стратегии проведения сопоставления, а простые сети — в качестве возможной вычис-лительной модели, то главная проблема, к которой мы обратим-ся в этом разделе, такова: *можно ли вычисления по методу ми-нимального отображения реализовать с помощью простой сети?* (Надо отметить, что такие простые сети в общем случае не экви-валентны универсальной вычислительной машине<sup>3,4</sup>.)

Вследствие дискретного, комбинаторного характера этой про-блемы перспективны проведение вычислений по методу минималь-ного отображения с помощью простой сети могут показаться сом-нительными. Однако в дальнейшем увидим, что если несколько изменить определение допустимых вариантов сопоставления, то вычисление соответствия на основе минимального отображения становится возможным. Вместо множества всех покрытий рас-смотрим подмножество локальных покрытий. У каждого элемен-та имеется  $N$  ближайших соседей, которые являются *начальными кандидатами* в допустимые варианты сопоставления<sup>3,5</sup>. Допусти-мым является тот вариант, при котором каждый элемент сопо-ставляется хотя бы с одним из своих начальных кандидатов.

Ищется тот допустимый вариант сопоставления, который минимизирует сумму  $\sum q(v_i)$ . В следующем разделе будет показано, что при такой формулировке оптимальный вариант сопоставления можно вычислить с помощью простого локального процесса. Определим также радиус вычисления, т. е. число соседей, с которыми должен быть соединен каждый процессор, чтобы вычисление по методу минимального отображения стало возможным. Для этого, как будет видно, достаточно, чтобы каждый процессор был соединен только со своими начальными кандидатами (т. е. чтобы  $r=N$ , где  $r$  — радиус вычисления;  $N$  — число начальных кандидатов). Это совсем не очевидно. Хотя каждый элемент и переходит при сопоставлении в один из своих начальных кандидатов, но вследствие общих требований согласованности и оптимальности (окончателное отображение должно быть покрытием минимальной стоимости) пара, назначенная данному элементу при сопоставлении, может зависеть от того, каким образом сопоставляются другие элементы, лежащие далеко за пределами множества его начальных кандидатов.

#### **3.4. Вычисление минимального отображения с помощью простой сети**

В этом и следующем разделах будет рассматриваться проблема вычисления минимального отображения с помощью простой локальной сети. Такое рассмотрение имеет формальный характер, и для удобства тех, кто хотел бы избежать математического анализа, основные результаты подытожены в начале разд. 3.4 и 3.5.

Основной результат этих разделов состоит в том, что минимальное отображение (а в силу этого и наиболее правдоподобный вариант соответствия) может быть вычислено по первоначальным степеням сродства с помощью простых локальных взаимодействий. Таким образом, в этом отношении минимальное отображение совместимо со схемой, описанной в гл. 2. Степени сродства (см. гл. 2) и стоимость  $q$  взаимно противоположны (например, степень сродства может быть равна  $C-q$ , где  $C$  — достаточно большая константа), так что минимизация  $q$  эквивалентна минимизации степеней сродства. Согласно этой точке зрения процесс установления соответствия, используемый зрительной системой человека, ведет к максимизации суммарного сродства и тем самым реализует наиболее правдоподобный вариант сопоставления.

Метод, с помощью которого простая сеть может вычислять минимальное отображение, представлен в разд. 3.4. Разработка этого метода включает два этапа: 1) переформулировка задачи вычисления минимального отображения в виде задачи линейного программирования, причем эквивалентность исходной задачи и формулировки в виде задачи линейного программирования обеспечивается теоремой из теории целочисленного программирования; на этом этапе задача сопоставления преобразуется из ди-

скретной задачи образования пар элементов в непрерывную задачу оптимизации; 2) использование метода, разработанного Эрроу, Гурвицем и Удзавой (1958), для решения полученной задачи линейного программирования с помощью простого локального процесса.

**Переформулировка задачи поиска минимального отображения в виде задачи линейного программирования.** Линейное программирование изучает задачи оптимизации линейных функций при наличии линейных ограничений. В векторных обозначениях задача линейного программирования имеет вид:

$$\begin{aligned} & \text{Минимизировать } \mathbf{c} \times \mathbf{x} & (3.4) \\ & \text{при условиях: } \mathbf{A}\mathbf{x} \geq \mathbf{b}; \\ & \mathbf{x} \geq 0, \end{aligned}$$

где  $A(a_{ij})$  — матрица размера  $m \times n$ ,  $\mathbf{x}$  и  $\mathbf{c}$  —  $n$ -мерные векторы, а  $\mathbf{b}$  —  $m$ -мерный вектор. Иначе говоря, требуется найти такой вектор  $\mathbf{x} = (x_1, \dots, x_n)$ , который будет минимизировать сумму  $\sum_i c_i x_i$  при наличии  $m$  ограничений на значения  $x_i$ . Ограничение с номером  $j$  имеет вид  $\sum_k a_{jk} x_k \geq b_j$  и  $x_i \geq 0$  для  $i = 1, \dots, n$ .

Для того чтобы перевести задачу нахождения минимального отображения на язык линейного программирования, введем переменные  $x_{ij}$ , где  $1 \leq i \leq n$ ,  $1 \leq j \leq m$  (если входных элементов  $n$ , а выходных —  $m$ ). Если  $i$ -й входной элемент при сопоставлении переходит в  $j$ -й выходной элемент, то  $x_{ij} = 1$ , в противном случае  $x_{ij} = 0$ . В случае покрытия  $\sum_i x_{ij} \geq 1$  для всех  $i$  и  $\sum_j x_{ij} \geq 1$  для всех  $j$ .

Поэтому сформулируем такую задачу линейного программирования:

$$\begin{aligned} & \text{Минимизировать } \sum x_{ij} q_{ij} & (3.5) \\ & \text{при условиях: } \sum_j x_{ij} \geq 1 \text{ для } 1 \leq i \leq n; \\ & \sum_i x_{ij} \geq 1 \text{ для } 1 \leq j \leq m; \\ & x_{ij} \geq 0 \text{ для } 1 \leq i \leq n, 1 \leq j \leq m. \end{aligned}$$

**Примечания:** 1) общее число переменных  $x_{ij}$  составляет  $nN$ , так как всего имеется  $n$  входных элементов, каждый из которых имеет  $N$  соседей-выходов; 2)  $q_{ij}$  представляет собой стоимость связи между  $i$ -м входным элементом и  $j$ -м выходным.

Эквивалентна ли эта задача линейного программирования исходной задаче поиска минимального отображения? Это было бы так при условии, что  $x_{ij}$  может принимать только два значения (т. е.  $x_{ij} = 1$  или  $x_{ij} = 0$ ). В рамках линейного программирования это дополнительное ограничение выразить невозможно, однако это, к счастью, является и лишним. Теорема из теории целочисленного программирования утверждает, что у этой задачи линейного программирования существует оптимальное решение.

ри котором все  $x_{ij}$  являются *целыми числами* [Гарфинкель и Земхойсер, 1972]. (Заметим, что матрица ограничений унимодальна.) Нетрудно проверить, что из нашего условия целочисленности следует, что в оптимальном решении единственными возможными для  $x_{ij}$  значениями являются 0 и 1. Следовательно, в предположении единственности решения любой алгоритм, решающий указанную задачу линейного программирования, заведомо решает исходную задачу нахождения минимального отображения. В тех редких случаях, когда оптимальное решение не единственно, имеются, по крайней мере, два различных оптимальных целочисленных решения, а также нецелочисленные решения. Сейчас предположим, что оптимальное решение единственно. Случай неединственного решения рассматривается в разд. 3.6.

Теперь опишем метод решения задачи линейного программирования, который может быть реализован с помощью простой сети.

**Вычисление минимального отображения в простой сети.** Метод оптимизации функций, в котором вычисления распределяются между простыми и локально связанными процессорами, был предложен в работе [Эрроу и др., 1958]. Он основан на теореме Куна — Такера [Кун и Такер, 1951], в которой устанавливается эквивалентность оптимальных решений задачи с ограничениями и седловых точек соответствующей функции Лагранжа.

Рассмотрим задачу максимизации функции  $f(x)$  при  $m$  ограничениях  $g_i(x) \geq 0, i=1, \dots, m$ . Функция Лагранжа, соответствующая данной задаче, определяется так:

$$L(x, u) = f(x) + \sum u_i g_i, \quad (3.6)$$

где  $x$  —  $n$ -мерный вектор;  $u$  —  $m$ -мерный вектор.

Неотрицательной седловой точкой функции Лагранжа является неотрицательная точка  $(x', u')$ , удовлетворяющая условию

$$L(x, u') \leq L(x', u') \leq L(x', u) \quad (3.7)$$

для всех  $x \geq 0; u \geq 0$ .

**Теорема** (Кун и Такер). Если 1)  $f(x)$  и  $g_i(x)$  вогнуты и 2) существует вектор  $x_0 \geq 0$  такой, что  $g_i(x_0) > 0$  ( $1 \leq i \leq m$ ), то вектор  $x'$  является решением задачи максимизации в том и только том случае, если существует вектор  $u'$  такой, что  $(x', u')$  является седловой точкой соответствующей функции Лагранжа. (Условия приведенной выше формулировки несколько отличаются от условий оригинальной теоремы Куна — Такера; доказательство см. в работе [Эрроу, Гурвиц и Удзава, гл. 3].) Заметим, что формулировка задачи установления соответствия в виде задачи линейного программирования удовлетворяет условиям теоремы Куна — Такера:  $f(x)$  и  $g_i(x)$  линейны, а в качестве  $x_0$  можно взять, к примеру, вектор, у которого  $x_{ij} = 2$  при всех  $i$  и  $j$ .

**Метод градиента функции Лагранжа.** Теорема Куна — Такера, являющаяся расширением классической теории множителей Лагранжа, переводит задачу оптимизации функции при ограниче-



ниях (задачу условной оптимизации) в задачу нахождения седловых точек соответствующей функции Лагранжа. Эрроу и другие (1958) исследовали возможности нахождения седловых точек с помощью градиентных методов. В методе градиента седловая точка функции  $L(\mathbf{x}, \mathbf{u})$  ищется путем сдвига в направлении локальных градиентов («подъем» по  $\mathbf{x}$ , «спуск» по  $\mathbf{u}$ ) с соблюдением условия неотрицательности переменных. Этот поиск формализуется с помощью дифференциальных уравнений Эрроу — Гурвица [Эрроу и другие, с. 118]:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{cases} 0, & \text{если } x_i = 0 \text{ и } Lx_i < 0; \\ Lx_i & \text{в противном случае;} \end{cases} \\ \dot{\mathbf{u}}(t) &= \begin{cases} 0, & \text{если } u_i = 0 \text{ и } Lu_i > 0; \\ -Lu_i & \text{в противном случае,} \end{cases} \end{aligned} \quad (3.8)$$

где  $Lx_i$  — частная производная функции Лагранжа по  $x_i$ ;  $Lu_i$  — по  $u_i$ . Можно осуществить приближение уравнений Эрроу—Гурвица с помощью следующих итераций:

$$\begin{aligned} x_i^{n+1} &= \max\{0, x_i^n + \rho Lx_i\}; \\ u_i^{n+1} &= \max\{0, u_i^n - \rho Lu_i\}, \end{aligned} \quad (3.9)$$

где  $\rho$  — выбранная длина шага. Если в этих формулах  $L$  является функцией Лагранжа того вида, который определен в теореме Куна — Такера, то метод носит название «наивного» метода, основанного на использовании функции Лагранжа\*. Главное обстоятельство, на которое следует обратить внимание, состоит в том, что вычисление обычного градиента в задаче нахождения минимального отображения является простым и локальным вычислением. Причина локальности заключается в том, что  $Lx_i$  и  $Lu_i$  выражены только через значения  $\mathbf{x}$  и  $\mathbf{u}$  на  $i$ -м процессоре и на  $N$  его непосредственных соседях. Точнее, функция Лагранжа равна

$$L(\mathbf{x}, \mathbf{u}) = -\sum q_{ij}x_{ij} + \sum_i u_i (\sum_j x_{ij} - 1) + \sum_j v_j (\sum_i x_{ij} - 1), \quad (3.10)$$

где  $u_i$  и  $v_j$  — множители Лагранжа,  $u_i$  соответствует входным элементам,  $v_j$  — выходным.

Если входных элементов  $n$ , а выходных  $m$ , то  $i = 1, \dots, n$  и  $j = 1, \dots, m$ . Производные принимают простой вид:  $Lx_{ij} = u_i + v_j - q_{ij}$ ;  $Lu_i = \sum_j x_{ij} - 1$ ;  $Lv_j = \sum_i x_{ij} - 1$ .

Так как производные локальны, то и процессы, задаваемые уравнениями (3.8) и (3.9), тоже локальны. Отметим, что для вычисления минимального отображения по этому методу каждый процессор должен вычислять лишь очень простую функцию соседей по сети. На рис. 3.2 приведен фрагмент сети, производящей вычисления в процессе установления соответствия, входные эле-

\* Этот метод не всегда гарантирует сходимость итераций к оптимальному решению (см. также ниже). — *Прим. ред.*

менты обозначены кружками, выходные — квадратами. Процессор  $u_i$  соответствует  $i$ -му входному элементу, процессор  $v_j$ — $j$ -му выходному элементу, процессор  $x_{ij}$  связи между  $i$ -м входным и  $j$ -м выходным элементами. В сети каждый процессор вычисляет простую функцию своих непосредственных соседей. Чтобы перейти

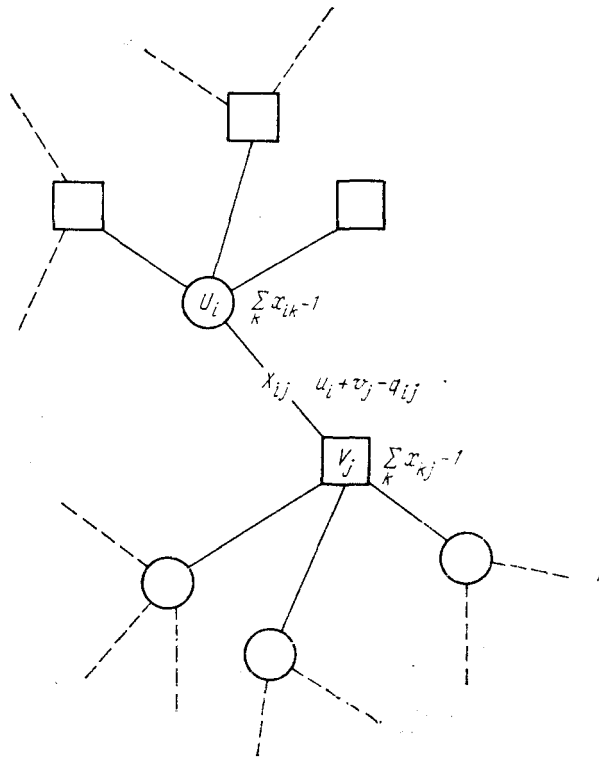


рис. 3.2. Часть простой сети, производящей вычисление минимального отображения. Рядом с процессорами указаны величины, которые они должны вычислять

из текущего состояния в следующее, процессоры должны вычислить такие величины (производные функции Лагранжа):

$x_{ij}$  вычисляет  $u_i + v_j - q_{ij}$ ;

$u_i$  вычисляет  $\sum_k x_{ik} - 1$ , где сумма берется по связям, исходящим из  $u_i$ ;

$v_j$  вычисляет  $\sum_k x_{kj} - 1$ , где сумма берется по связям, входящим в  $v_j$ .

**Сходимость.** Говорят, что метод Эрроу — Гурвица сходится к решению, если  $(\mathbf{x}(t), \mathbf{u}(t))$  стремится к седловой точке функции  $L(\mathbf{x}, \mathbf{u})$  при  $t \rightarrow \infty$ . Обычный градиентный метод в том виде, как он описан выше, не обязательно сходится к решению. Если  $L(\mathbf{x},$

и) линейна по обоим аргументам, то решение может вместо сходимости войти в предельный цикл [Эрроу и другие, 1958, гл. 6]. Есть, однако, некоторые способы улучшить поведение решения. Можно, например, использовать тот факт, что введением пороговой операции окончательное значение каждого  $x_{ij}$  можно свести к 0 или 1: та связь, для которой значение  $x_{ij}$  превышает определенный порог, становится частью покрытия, тогда как связь, для которой  $x_{ij}$  лежит ниже этого порога, устраняется. Такого рода операция может также служить и для ускорения процесса, ибо (3.9), как правило, сходится медленно<sup>3,6</sup>.

Имеются и общие способы видоизменения метода, основанного на использовании обычной функции Лагранжа, при котором обеспечивается глобальная сходимость уравнений Эрроу — Гурвица к седловой точке функции Лагранжа (как видоизмененной, так и исходной) и, следовательно, к оптимальному решению. Эту модифицированную функцию Лагранжа  $LM$  можно определить, например, так [Эрроу и другие, 1958, с. 137]:

$$LM(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \sum u_i \psi_i[g_i(\mathbf{x})], \quad (3.11)$$

где  $\psi_i$  — строго возрастающие, строго вогнутые аналитические функции, подчиненные условию  $\psi_i(0) = 0$ . (Примером такой функции является  $\psi(z) = 1 - e^{-rz}$  при  $r > 0$ .) Если к модифицированной функции Лагранжа применить непрерывный процесс (3.8), то для вогнутых  $f$  и  $g$ , в том числе и для линейного случая, сходимость обеспечена. Отметим, что градиентный метод, примененный к модифицированной функции Лагранжа, по-прежнему приводит к локальному вычислению, похожему на вычисление в случае «наивного» градиентного метода (хотя и не столь же простому).

Если к этой модифицированной функции Лагранжа применяется итеративная процедура (3.9), то итерации, как правило, тоже сходятся к решению<sup>3,7</sup>. Более того, можно еще и так видоизменить исходную функцию Лагранжа, чтобы гарантировать глобальную сходимость итеративной процедуры к решению (при достаточно малой длине шага  $\rho$ ) с сохранением свойства локальности процедуры. Как и прежде, градиенты  $LM_x$  и  $LM_y$ , подсчитанные для  $i$ -й компоненты, зависят только от  $i$ -го процессора и  $N$  его непосредственных соседей.

**Выводы.** Минимальное отображение, которое при принятии предположения о независимости одновременно является и оптимальным вариантом сопоставления, можно определить с помощью простого локального вычисления. Можно представить себе сеть, составленную из простых процессоров, которая воспринимает два «моментальных снимка» движущихся элементов и с помощью локальных взаимодействий находит наиболее правдоподобное соответствие между ними. Этот метод можно применить и к другим задачам условной оптимизации, в которых вычисления проводятся с помощью локальных процессов [Ульман, 1978с]. Следует отметить, что в приведенной схеме используется то упрощение, будто связь между элементами может находиться только в

вух состояниях: либо есть ( $x_{ij}=1$ ), либо нет ( $x_{ij}=0$ ). Здесь не было предпринято попытки связать с данным вариантом сопоставления степень его «проявления», как это было сделано в гл. 2. Степень проявления можно ввести в эту схему разными способами. Например, хотя  $x_{ij}$  и сходится только лишь к 0 или 1, но до того момента, пока эта сходимость не осуществится, он может принимать и промежуточные между 0 и 1 значения, которые можно использовать в качестве отражающих степень проявления данного варианта сопоставления.

### 3.5. Предпочтительность взаимно-однозначных отображений

Метод минимального отображения в приведенном виде не «штрафует» сопоставления за то, что они отклоняются от взаимно-однозначности. Такое упрощение не удовлетворяет нас ни с теоретической, ни с практической точки зрения.

С теоретической точки зрения, расщепления и слияния элементов на реальных изображениях маловероятны, но не исключены. Если, например, получится так, что на каком-то изображении один элемент загородит собой другой, то в следующий момент тот единственный элемент может расщепиться и превратиться в два элемента. Такой случай довольно редок, и потому тех отображений, при которых происходят расщепления и слияния, надо избегать, но избегать не любой ценой. Пусть  $\delta$  — вероятность загоразивания или появления любого другого источника расщеплений и слияний на изображениях. Тогда вероятность простого расщепления (при котором на изображении входной элемент расщепляется и дает два выходных элемента) или простого слияния (два выходных элемента сходятся на один и тот же выходной элемент) равняется  $\delta$ . Вероятность того, что элемент обладает тремя связями («двойное загоразивание»), есть  $\delta^2$ . В общем случае вероятность расщепления с числом связей  $s+1$  составляет  $\delta^s$ , а вероятность слияния с числом связей  $f+1$  —  $\delta^f$ . Вероятность варианта сопоставления, содержащего  $k$  расщеплений с  $s_1+1, \dots, s_k+1$  связями и  $n$  слияний с  $f_1+1, \dots, f_n+1$  связями, дается выражением

$$\prod p(v_i) \delta^{s_1} \dots \delta^{f_n}. \quad (3.12)$$

Взяв от выражения (3.12) логарифм со знаком минус, получим, что «стоимость» этого сопоставления равна (здесь  $\sigma = -\log \delta$ )

$$\sum q(v_i) + \sigma(\sum s_i + \sum f_j), \quad i=1, \dots, k; \quad j=1, \dots, n. \quad (3.13)$$

Для нахождения оптимального сопоставления минимизируется (3.13), а не (3.3). Чем больше в последнем выражении значение  $\delta$  (т. е. чем меньше вероятность расщеплений и слияний), тем большее предпочтение будет отдаваться взаимно-однозначным отображениям.

Назначение дополнительного штрафа за расщепления и слияния обусловлено также и экспериментальными результатами. Пример этого дан на рис. 3.3. Вариант сопоставления на рис. 3.3а

$(A1 \rightarrow B1 \leftarrow A2, B2 \leftarrow A3 \rightarrow B3)$  минимизирует  $\Sigma q_i$  (что остается верным и при большом МСИ, см. разд. 3.6), однако наше восприятие предпочитает взаимно-однозначное сопоставление, представленное на рис. 3.3б ( $A1 \rightarrow B1, A2 \rightarrow B2, A3 \rightarrow B3$ ).

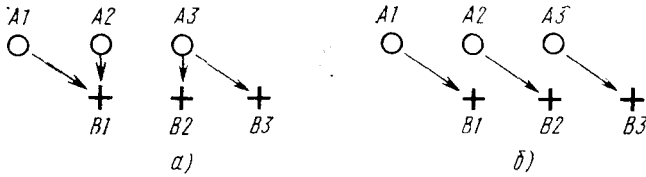


Рис. 3.3. Предпочтительность взаимно-однозначных отображений: вариант сопоставления (б) предпочтительнее, хотя общая стоимость варианта (а) ниже.

В этом разделе устанавливается неожиданный, быть может, результат: метод минимального отображения очень просто «настроить» так, чтобы устранить любое желаемое число расщеплений и слияний. Единственное видоизменение, которое для этого требуется, состоит в прибавлении одной константы к функции стоимости  $q$ . Покажем, что значение константы составляет  $2\sigma$ , где, как и прежде,  $\sigma = -\log \delta$ .

**Модифицированный метод минимального отображения.** Предположим, что теперь надо минимизировать не  $\Sigma q_{ij}x_{ij}$ , а сумму  $\Sigma q_{ij}x_{ij} + k\Sigma x_{ij}$  (или, что то же самое,  $\Sigma (k + q_{ij})x_{ij}$ ). Как и прежде,  $x_{ij}$  в оптимальном решении будет принимать только два возможных значения, следовательно, «штрафная» функция  $k\Sigma x_{ij}$  будет равна просто  $k$ , умноженному на общее число связей в данном варианте сопоставления. Увеличивая  $k$ , можно добиться преимущества для отображений с меньшим числом связей. Более того, следующее предложение показывает, что выбором подходящего значения  $k$  можно минимизировать нужную сумму в выражении (3.13).

**Предложение.** Минимизация суммы  $\Sigma q_{ij}x_{ij} + 2\sigma\Sigma x_{ij}$  (при наличии обычных ограничений  $\Sigma x_{ij} \geq 1$ ) эквивалентна минимизации штрафной суммы в (3.13) по множеству всех покрытий.

**Доказательство.** Заметим для начала, что наличие цепей из соответствующих друг другу элементов исключено. Действительно, рассмотрим цепь  $A1 \rightarrow B1 \leftarrow A2 \rightarrow B2$ . Связь  $B1 \leftarrow A2$  можно удалить, не нарушая при этом ограничений, следовательно, цепь не может быть частью оптимального решения. Пусть  $m$  — число однозначных связей в данном варианте сопоставления. Тогда общее число связей в этом варианте равно

$$\Sigma x_{ij} = m + \Sigma (s_i + 1) + \Sigma (f_j + 1), \quad (3.14)$$

где  $i$  пробегает номера расщеплений,  $j$  — слияний. Число входных элементов  $I$  дается выражением

$$I = m + \Sigma |s| + \Sigma (f_j + 1), \quad (3.15)$$

де  $|s|$  — общее число расщеплений. Число выходных элементов  $O$  дается выражением

$$O = m + \sum (s_i + 1) + |f|, \quad (3.16)$$

де  $|f|$  — общее число слияний. Вычтем теперь  $\sigma(I+O)$  из нашей целевой функции. Результат не зависит от варианта сопоставления, следовательно, не изменяет задачи оптимизации (т. е. вариант сопоставления минимизирует сумму  $\sum q_{ij} + 2\sigma \sum x_{ij}$  в том и только в том случае, если он также минимизирует  $\sum q_{ij}x_{ij} + 2\sigma \sum x_{ij} - \sigma(I+O)$ . После подстановки выражений для  $\sum x_{ij}$ ,  $I$  и  $O$  штрафная функция  $2\sigma \sum x_{ij} - \sigma(I+O)$  принимает вид

$$\sigma(\sum s_i + \sum f_j). \quad (3.17)$$

Значит, минимизация суммы  $\sum q_{ij}x_{ij} + 2\sigma \sum x_{ij}$  эквивалентна минимизации суммы  $\sum q_{ij}x_{ij} + \sigma(\sum s_i + \sum f_j)$ . Так как  $x_{ij}$  принимают только два возможных значения (и связаны условием  $\sum x_{ij} \geq 1$ ), то последняя минимизация эквивалентна минимизации  $\sum q_{ij} + \sigma(\sum s_i + \sum f_j)$  на множестве всех покрытий  $\blacklozenge$ .

Подводя итоги, можно сказать, что штрафную сумму в (3.13) можно оптимизировать, не затрагивая выражения, описанного в разд. 3.4. В стоимость  $q_{ij}$  можно просто включить константу  $k$ , так что оптимальное решение по-прежнему ищется путем минимизации  $\sum q_{ij}x_{ij}$ . Ответственность за оказание предпочтения взаимно-однозначным отображениям можно распределить по всей сети, просто прибавляя  $2\sigma$  к функции стоимости  $q(v)$ . Вычисление при этом остается простым и локальным, в то же время проявляя в требуемой мере предпочтение взаимно-однозначным отображениям.

Чтобы оценить значение этого раздела, рассмотрим, к примеру, сопоставление  $n$  входных элементов с  $n+1$  выходными. Предположим, что штраф достаточно велик, так что сопоставление будет настолько взаимно-однозначным, насколько это возможно, т. е. расщепление будет только одно. В целом сопоставление должно содержать единственное расщепление, и результирующий вариант сопоставления должен минимизировать общую стоимость. Если какой-то элемент «решил» расщепиться, то он должен «удостовериться», что больше ни один элемент не расщепился. Кроме того, расщепиться должен именно нужный элемент, ибо в противном случае получим неоптимальность функции стоимости. Поэтому создается впечатление, что проведение правильного сопоставления может потребовать осуществления сложных взаимодействий, обмена информацией между удаленными друг от друга элементами или какого-то глобального процесса управления, координирующего все соединения элементов в пары. Однако из предыдущих двух разделов вытекает, что ничего подобного не нужно. Оптимального варианта сопоставления с единственным расщеплением можно достигнуть, используя только простые локальные взаимодействия.

Таким образом, можно сделать вывод, что минимальное отображение, обладающее соответствующей настройкой на взаимнооднозначные отображения, можно вычислить с помощью локальных взаимодействий, зная значения исходных степеней родства. Следовательно, минимальное отображение является работоспособной схемой, перспективной для объяснения того процесса установления соответствия, который используется зрительной системой человека. Однако описанная сеть не предлагается нами в качестве точной модели. Необходимо делать различие между моделью и изучением основных принципов построения системы с точки зрения вычислений. В точной модели дается формальное описание системы и утверждается, что элементы модели и элементы моделируемой системы изоморфны. Наша главная цель заключается не в моделировании процесса установления соответствия, а в исследовании некоторых вычислительных аспектов этой проблемы. Основные результаты исследования состоят в том, что при определенных условиях минимальное отображение является оптимальным, и с помощью локального процесса в соответствии со схемой, описанной в гл. 2, его можно вычислить через исходные степени родства. С учетом этой точки зрения роль исследованных в гл. 2 взаимодействий заключается в увеличении общего родства, что в результате дает минимальное отображение (или, по крайней мере, его приближение) и тем самым проводит наиболее правдоподобное сопоставление между участвующими в нем опорными символами. В следующем разделе будет продолжено изучение черт сходства между минимальным отображением и процессом установления соответствия, используемого зрительной системой человека.

### 3.6. Свойства минимального отображения

До сих пор оптимальная стратегия установления соответствия характеризовалась нами с помощью некоторого математического условия, а именно через минимизацию функции стоимости на множестве всех локальных покрытий. В этом разделе обратимся к изучению некоторых свойств минимального отображения и сравним их (там, где это возможно) со свойствами того соответствия, которое устанавливается зрительной системой человека.

**Минимизация суммарного пути.** Можно ожидать, что по мере возрастания межстимульного интервала между последовательными кадрами минимальное отображение будет минимизировать суммарный путь, проходимый элементами изображения. Это значит, что если  $d_i$  — путь, пройденный  $i$ -м входным элементом, то минимальное отображение будет минимизировать сумму  $\sum d_i$  на множестве всех допустимых вариантов сопоставления.

В случае малых скоростей  $p(v)$  можно представить как  $p_0 + kv$ , где  $p_0 = p(0)$ , а  $k$  — производная  $p(v)$  в точке 0 (далее увидим, что  $k < 0$ ). Для малых  $v$  имеет место следующее приближение:

$$q(v) = -\log p(v) = -\log(p_0 + kv) = \alpha + \beta v, \quad (3.18)$$

где  $\alpha = -\log p_0$ ;  $\beta = -k/p_0$ .

Если, к примеру, рассматриваются взаимно-однозначные отображения (как это будет в дальнейшем), то вклад, вносимый  $\alpha$ , будет постоянен и предпочтительный вариант сопоставления будет минимизировать сумму скоростей  $\sum v_i$ . В случае двух кадров при большом МСИ между ними это эквивалентно также минимизации суммарного пути  $\sum d_i$  (так как  $d_i = v_i \times \text{МСИ}$ )<sup>3,8</sup>.

**Правило непересекающихся траекторий.** Как уже отмечалось, при кажущемся движении траектории элементов редко пересекаются друг с другом [Колерс, 1972; Эттлив, 1974; Нэвон, 1976]. Если  $A1$  и  $A2$  предъявляются при условиях кажущегося движения  $B1$  и  $B2$  в такой конфигурации, что траектории движения  $A1 \rightarrow B2$  и  $A2 \rightarrow B1$  пересекаются между собой, а траектории  $A1 \rightarrow B1$  и  $A2 \rightarrow B2$  не пересекаются (рис. 3.4), то второй вариант сопоставления будет предпочтительнее (если МСИ не слишком мал).

Правило непересекающихся траекторий содержится (частично, но крайней мере) в принципе минимального пути. Из неравенства треугольника, примененного к рис. 3.4, вытекает, что  $d_1 + d_2 < c_1 + c_2$ . Это значит, что непересекающиеся траектории всегда минимизируют суммарный путь и, следовательно, при большом МСИ минимизируют и сумму  $\sum q_i$ .

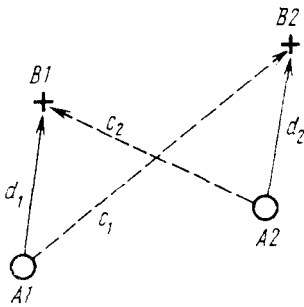


Рис. 3.4. Непересекающиеся траектории; суммарный путь непересекающихся траекторий всегда меньше, чем у пересекающихся

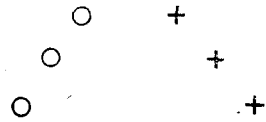


Рис. 3.5. Групповые опорные символы; точки могут быть сгруппированы в отрезки до установления соответствия

В работе [Колерс, 1972, с. 77] приведены интересные примеры, показывающие, что некоторые эксперименты с непересекающимися траекториями нельзя объяснить на основе одного только суммарного пути. Делается вывод, что зрительная система «не выносит» пересекающиеся траектории (кроме особых случаев, относящихся к движению вглубь). Нам представляется, однако, что вовсе не обязательно делать предположение о том, что зрительная система налагает на пересекающиеся траектории явный запрет. Во-первых, как этого и следовало ожидать от минимального отображения, можно получить пересекающиеся траектории, манипулируя значениями степеней сродства между входящими в изображение элементами (т. е. если на рис. 3.4 сродство  $A2$  и  $B1$  достаточно велико). Во-вторых, иногда надо учитывать также и эффекты группировки. В подразд. 2.4.2 уже упоминалось о возможности группирования элементарных опорных символов в опор-



ные символы более высокого уровня. Например, точки на рис. 3.5 можно сгруппировать в отрезки соответствующих длины и ориентации. Тогда соответствие можно устанавливать либо между отдельными точками, либо между двумя опорными символами групп. При этом оба варианта являются конкурирующими друг с другом и «победа» одного из них зависит от временных параметров и параметров сродства. Это дополнительное состязание можно включить в схему минимального отображения, изложенную в разд. 3.4 и 3.5. Такое включение представляет собой крайне важное расширение схемы минимального отображения. Но поскольку явление группирования изучено недостаточно, то это расширение в дальнейшем обсуждаться не будет. Здесь же достаточно заметить, что, основываясь на рассмотрении разнообразных примеров [например, примеры 22 и 23 в работе Колерса, 1972], можно сказать, что длина траектории и группировка, по-видимому, достаточны для объяснения причин отсутствия пересечений траекторий<sup>3,9</sup>.

**Обнаружение сдвигов.** Предположим, что с набора элементов, движущихся параллельно друг другу, сделаны два «моментальных снимка»:  $S1$  и  $S2$ . Назовем такое движение *сдвигом* элементов. Судя по всему зрительная система обладает способностью обнаруживать сдвиги: если  $S1$  и  $S2$  предъявлять один за другим, то, как правило, воспринимается движение типа сдвига (если МСИ не слишком мал). Это остается в силе даже тогда, когда среднее расстояние, пройденное элементами в промежутке между двумя снимками, значительно больше расстояния между самими элементами; в этом случае практически ни один элемент не сопоставляется со своими ближайшими соседями<sup>3,10</sup>.

Такая способность к обнаружению сдвига заслуживает более пристального изучения, поскольку она, на первый взгляд, не согласуется с гипотезой независимости, выдвинутой в разд. 3.2. Может показаться, что способность к обнаружению сдвига свидетельствует о том, что каждый элемент «предпочитает» при сопоставлении выбрать себе пару в том направлении, которое согласуется с направлением движения соседних элементов. Однако гипотеза независимости исключает те взаимодействия, которые основаны на сходстве направлений. Явление обнаружения сдвига может также указывать и на существование неких глобальных средств измерения, которые не относятся ни к одному отдельно взятому процессору в той простой сети, которая обсуждалась в разд. 3.4. С помощью глобальных средств измерения можно выявлять преобладающее направление и затем оказывать влияние на варианты сопоставления, назначаемые отдельным элементам. Но такое предположение о взаимодействии локальных процессов с глобальными идет вразрез с нашей простой моделью сети. Следовательно, явление обнаружения сдвига приводит к возникновению следующей трудности. В модели простой сети установление соответствия между наборами элементов полностью управляется локальными взаимодействиями. В соответствии с гипотезой неза-

зависимости в эти локальные взаимодействия не включаются положительные взаимодействия между сопоставлениями в сходных направлениях. Тем не менее если общее направление все же существует, то оно, скорее всего, оказывает влияние на процесс установления соответствия, о чем свидетельствует явление обнаружения сдвига.

В целях преодоления этой трудности обратимся к изучению явления обнаружения сдвига с точки зрения минимального отображения. Вывод, к которому мы придем, состоит в том, что обнаружение сдвига не противоречит ни предположению о независимости, ни модели простой сети. На самом деле оно служит их подтверждением, поскольку, как будет видно, минимальное отображение фактически предполагает наличие обнаружения сдвига.

Напомним, что  $S_2$  получается из  $S_1$  при сдвиге всех элементов вдоль некоторого направления. Правильное сопоставление между  $S_1$  и  $S_2$  — это такое сопоставление, при котором каждый элемент из  $S_1$  переходит в свой, подвергшийся сдвигу образ на изображении  $S_2$ . Теперь докажем следующее утверждение.

**Утверждение** (лемма об обнаружении сдвига). Правильное сопоставление минимизирует суммарный путь  $\sum d_i$  на множестве всех взаимно-однозначных отображений.

**Доказательство.** Обозначим через  $(x_i, y_i)$  положение  $i$ -го входного элемента в плоскости изображения, а через  $(x'_i, y'_i)$  его положение на втором кадре. Если ось  $X$  выбрана совпадающей с направлением сдвига, то  $y'_i = y_i$ , а  $x'_i \geq x_i$ . Сопоставление кадров друг с другом представляет собой функцию  $m$ , которая каждому входному элементу назначает некоторый выходной. Таким образом,  $j = m(i)$  означает, что  $i$ -й входной элемент переходит в  $j$ -й выходной.

Суммарный путь  $D_c$  при правильном сопоставлении дается выражением

$$D_c = \sum (x'_i - x_i) = \sum x'_i - \sum x_i. \quad (3.19)$$

Для другого варианта сопоставления  $m$  суммарный путь  $D_m$  дается выражением

$$D_m = \sum ((x'_j - x_i)^2 + (y'_j - y_i)^2)^{1/2}, \quad i = 1, \dots, n; \quad j = m(i). \quad (3.20)$$

Теперь можно сравнить между собой суммарный путь для вариантов сопоставления  $m$  и  $c$ :

$$\begin{aligned} D_m &\geq \sum |x'_j - x_i| \geq \sum (x'_j - x_i) = \\ &= \sum x'_j - \sum x_i = D_c, \quad i = 1, \dots, n; \quad j = m(i). \end{aligned} \quad (3.21)$$

Поскольку  $D_m \geq D_c$ , то  $D_c$  минимально, а правильное сопоставление оптимально  $\blacklozenge$ .

Можно также показать, что в случае, когда  $m$  не является сдвигом (т. е. когда не выполняется  $y'_j = y_i, x'_j \geq x_i$ ),  $D_m$  строго больше  $D_c$ . Поэтому оптимальный вариант сопоставления будет оптимальным всегда, за исключением некоторых особых ситуаций.

Лемму об обнаружении сдвига можно доказать также и для случая радиального движения. Предположим, что в плоскости изображения каждый элемент движется по прямой, соединяющей его с некоторой фиксированной точкой  $o$ . (Такое радиальное движение может возникнуть при наблюдении приближающегося объекта или перспективной проекции чистого параллельного переноса в трех измерениях.) Тогда правильное радиальное соответствие минимизирует суммарный путь.

**Доказательство.** Пусть  $o$  — начало координат, а положение каждого элемента будем описывать его полярными координатами  $(r, \theta)$ . Если  $d_{ij}$  — расстояние между входным элементом  $i$  и выходным  $j$ , то  $d_{ij} \geq |r_i - r_j|$ .

Для данного варианта сопоставления  $m$  имеем

$$D_m = \sum d_{ij} \geq \sum |r_i - r_j| \geq \sum (r_i - r_j) = D_r, \quad i=1, \dots, n; j=m(i), \quad (3.22)$$

где  $D_r$  — суммарный путь при правильном (радиальном) соответствии  $\blacklozenge$ .

**Примечание:** с точки зрения обнаружения сдвига вращательное движение отличается от параллельного переноса и радиального движения. Если  $S2$  получено из  $S1$  в результате поворота на достаточно большой угол, то минимальное отображение не всегда будет соответствовать такому повороту. Можно ожидать, что в соответствующих условиях дискретно предъявляемое вращательное движение разрушится при меньших средних значениях смещения, чем в случае параллельного переноса и вращения<sup>3,11</sup>. Следует также отметить, что в восприятие согласованного движения могут вносить свой вклад и дополнительные источники. Если смещение малое, то оно может быть обнаружено процессом с малой зоной действия, о котором говорилось в разд. 1.2. Этот процесс, видимо, пригоден даже тогда, когда число опорных символов становится столь велико, что исключает установление соответствия между всеми отдельно взятыми опорными символами.

**Еще раз о гипотезе независимости.** Оптимальная стратегия установления соответствия была разработана для элементов, движущихся независимо друг от друга. Предположение о независимости можно оспорить тем, что близко расположенные элементы изображения с высокой вероятностью движутся в сходных направлениях. Можно, следовательно, утверждать, что если существует локально параллельное сопоставление (т. е. сопоставление, при котором пространственно близкие элементы движутся почти параллельно), то оно должно быть предпочтительным. Хотя в этом доводе и есть, вероятно, доля правды, однако анализ обнаружения сдвига показывает, что явное включение в схему предпочтения такого рода является излишним, ибо параллельные движения тоже минимизируют  $\sum d_i$ . Таким образом, метод минимального отображения приемлем независимо от того, является движение элементов на самом деле независимым или нет. Следует также отметить, что в методе минимального отображения требуется определять только расстояния между элементами, а не направле-

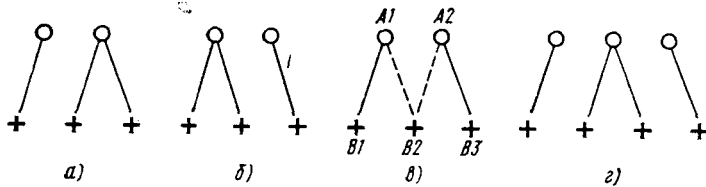


рис. 3.6. Симметричные отображения: оптимальные (а, б); оптимальное, но не целое и неустойчивое (в); оптимальное целое и устойчивое (г)

тия; это свойство может быть полезно с точки зрения экономической реализации. Поскольку сцена может содержать много элементов, то вычисление оптимального соответствия на основе минимального числа параметров и при минимальном числе взаимодействий может дать существенные преимущества.

**Симметрия.** Одним из свойств процесса установления соответствия, используемого зрительной системой человека, является предпочтение, отдаваемое симметричному движению при равенстве прочих, более важных параметров» [Этнвив, 1974, с. 118]. Такого свойства симметричности следует ожидать в любой простой локальной сети типа описанной в разд. 3.4. Кроме того, если во входной конфигурации имеется симметрия, то должен существовать и симметричный вариант оптимального сопоставления. Симметрию можно определить как перестановку  $\pi$ , которая «не изменяет задачи». Иначе говоря, если  $q_{ij}$  — стоимость связи между входным элементом  $i$  и выходным  $j$ , а  $q'_{ij}$  — стоимость связи между  $\pi(i)$  и  $\pi(j)$ , то для всех  $i$  и  $j$  должно быть  $q_{ij} = q'_{ij}$ . Если такая симметрия существует, то: 1) существует симметричный вариант оптимального сопоставления, т. е. сопоставления, при котором  $x_{ij} = x'_{ij}$ . Это происходит потому, что если  $(x_{ij})$  (последовательность нулей и единиц) является оптимальным решением, то это же верно и для  $(x'_{ij})$ . Решение же  $(y_{ij})$ , определенное соотношением  $y_{ij} = (x_{ij} + x'_{ij})/2$ , также оптимально и уже симметрично; 2) итеративная процедура (3.9) будет сходиться к симметричному оптимальному решению (если оно является допустимым вариантом сопоставления). Так как вход симметричен, то и результат первого шага итерации симметричен. Так как все процессоры одинаковы, то результат следующего шага (и по индукции всех шагов) тоже будет симметричным.

Симметричные конфигурации можно разделить на две категории: целые и нецелые\*. На рис. 3.6 приведен пример нецелой симметричной конфигурации, на рис. 3.6а показано одно оптимальное отображение, на рис. 3.6б — другое. Отображение на рис. 3.6в представляет собой комбинацию двух предыдущих и является одновременно оптимальным и симметричным. Как было отмечено в разд. 3.4, в том случае, когда оптимальное решение

\* Это определение (по-английски integer/non-integer) относится сразу к обоим предъявляемым изображениям. — Прим. ред.

не единственно (и только в этом случае), существуют также и нецелые оптимальные решения, примером которых и является рис. 3.6в. Можно ожидать, что отображение на рис. 3.6в будет неустойчиво, так как оно основывается на точном равенстве расстояний между  $A1$  и  $B2$ , с одной стороны, и  $A2$  и  $B2$  — с другой. Любое отклонение этих расстояний от строгого равенства друг другу приведет к тому, что оптимальной будет конфигурация либо на рис. 3.6а, либо на рис. 3.6б. Неудивительно поэтому, что восприятие этой конфигурации неустойчиво и переменчиво [Колерс, 1972; Эттнив, 1974; см. разд. 2.4 настоящей работы].

На рис. 3.6г показаны оптимальное, симметричное и целое отображения. Такие отображения в противоположность нецелым воспринимаются нами как устойчивые. Устойчивость не может быть выведена только из минимального отображения, так как она зависит и от свойства алгоритма, реализующего этот метод. Можно проверить, что если  $n$  элементов, расположенных в строку, показываются поочередно со строкой из  $n+1$  элемента, то при четном  $n$  симметричное решение является нецелым, а при нечетном  $n$  существует симметричное оптимальное целое решение. Поэтому есть основания полагать, что в первом случае воспринимаемое сопоставление будет неустойчивым и асимметричным, а во втором — устойчивым и симметричным. Это предсказание согласуется с данными работ [Колерс, 1972; Эттнив, 1974].

**Симметрия в порядке предъявления.** Если два кадра  $f1$  и  $f2$  показываются в условиях кажущегося движения, то воспринимаемое соответствие не зависит от порядка предъявления. Это означает, что переходы элементов друг в друга остаются теми же независимо от того, показывается  $f1$  до или после  $f2$ . Эта же симметрия в равной мере свойственна и процессу установления соответствия на основе минимального отображения: входные и выходные элементы меняются ролями, а оптимальное решение задачи выбора сопоставления остается инвариантным.

**Свойство минимального покрытия.** Минимальное отображение является минимальным покрытием в том смысле, что оно не содержит лишних связей. Удаление из варианта сопоставления любой связи приведет к тому, что один из входных или выходных элементов окажется без покрытия (т. е. без парного элемента). Это свойство подразумевает наличие явлений расщепления и слияния, обсуждавшихся в разд. 2.4. На рис. 3.7 дан пример конкуренции при расщеплении. На рис. 3.7а вслед за предъявлением элемента  $A1$  предъявляется пара элементов  $B1$  и  $B2$ , распо-

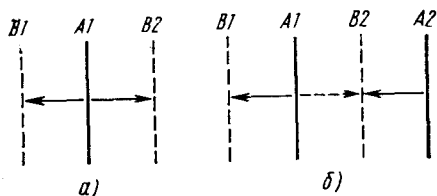


Рис. 3.7. Свойство минимального покрытия:  $A1$  расщепляется (а);  $A1$  не будет расщепляться, так как связь  $A1 \rightarrow B1$  была бы тогда излишней (б)

енных по бокам от  $A1$ . Элемент  $A1$  воспринимается расщепляющимся и сопоставляется как с  $B1$ , так и с  $B2$ . На рис. 3.76 на первом кадре добавлен еще один элемент  $A2$ . В результате получается соответствие  $A1 \rightarrow B1$ ,  $A2 \rightarrow B2$ , а связь  $A1 \rightarrow B2$  исчезает. Все выглядит так, словно  $A2$ , перехватив  $B2$ , соревнуется с  $A1$  и не дает ему сопоставиться с  $B2$ . При минимальном отображении три связи ( $A1 \rightarrow B1$ ,  $A2 \rightarrow B2$ ,  $A1 \rightarrow B2$ ) не могут существовать одновременно, так как тогда это отображение не будет минимальным покрытием (ведь  $A1 \rightarrow B2$  можно удалить). Аналогичным образом Эттнев (1974) описал конфигурацию, в которой при кажущемся снижении число связей остается на уровне минимума, требуемого для обеспечения каждого элемента партнером. Описывая это свойство минимального покрытия, а также такие свойства, как симметрия и непересечение траекторий, Эттнев (1974) замечает: Создается впечатление, что эта система проявляет способности предвидения, и поэтому возникает сильное искушение призвать некое «*deus ex machina*»\*, какую-то рационально организованную систему управления более высокого уровня, которая устроила бы все нужным образом [Эттнев, 1974, с. 116].

В результате проведенного в предыдущих разделах обсуждения установлено, что в такого рода глобальном планировании нет необходимости. Всеми указанными свойствами может обладать и простой локальный процесс, и можно ожидать, что он их проявит на деле в результате вычисления минимального отображения.

**Монотонная зависимость от частоты регистрации.** Грубо говоря, если отображение, минимизирующее  $\Sigma d_i$ , при разнесении кадров на временной интервал  $t$  дает правильное установление соответствия, то и для всех  $t' < t$  этот вариант сопоставления будет тоже правильным.

Для доказательства этого утверждения будем считать, что элементы движутся вдоль прямых линий (такое предположение будет справедливо для малых временных интервалов). Пусть даны два мгновенных изображения движущихся элементов, разделенные временным интервалом  $t$ . Предположим, что правильное сопоставление (т. е. тот вариант сопоставления, при котором каждый элемент переходит в свой образ через время  $t$ ) минимизирует  $\Sigma d_i$ . Тогда для всех  $t' < t$  правильное соответствие также будет минимизировать  $\Sigma d_i$ .

**Доказательство.** Пусть  $v_i$  — скорость  $i$ -го элемента;  $\sigma$  — суммарный путь  $\Sigma d_i$  при правильном сопоставлении в момент  $t$ , а  $\sigma'$  — суммарный путь в момент  $t'$ . Пусть  $\mu'$  — суммарный путь  $\Sigma d_i$  некоторого взаимно-однозначного сопоставления  $m'$  в момент  $t'$ . Докажем, что  $\sigma' \leq \mu'$ . Вариант сопоставления  $m$  в момент  $t$  можно получить из варианта сопоставления  $m'$  в момент  $t'$  следующим образом: если точка  $x$  под действием  $m'$  переходит в точку

\* «Божество из машины» (лат.) — вариант развязки классической античной трагедии, когда все коллизии разрешались небожителем, внезапно и без всякой связи с предыдущим, появившимся на сцене. — *Прим. пер.*

$y(t')$  (т. е. в точку  $y$  в момент  $t'$ ), то  $m$  получается благодаря связыванию  $x$  с  $y(t)$ . На рис. 3.8 представлены различные варианты сопоставления. Правильным является сопоставление  $y(0) \rightarrow y(t') \rightarrow y(t)$ . При  $m': x(0) \rightarrow y(t')$ , а при  $m: x(0) \rightarrow y(t)$ . Через  $\mu$  будем обозначать суммарный путь в новом варианте сопоставления  $m$ . Из допущения о том, что правильное сопоставление минимизирует сумму  $\sum d_i$  в момент  $t$ , следует, что  $\sigma \leq \mu$ . Тогда для доказательства нашего утверждения достаточно показать, что  $\sigma - \sigma' \geq \mu - \mu'$ . Для этого обратимся к рис. 3.8. Вклад элемента  $y$  в  $\sigma$  составляет  $v_y t$ , а в  $\sigma' - v_y t'$ . Разность этих вкладов составляет  $v_y(t - t')$  и на рис. 3.8 представлена отрезком  $r$ . При сопоставлении  $m: x \rightarrow y(t)$ , а при  $m': x \rightarrow y(t')$ . На рис. 3.8 разность между этими двумя вкладами в значение  $y$  составляет  $d - d'$ , а по неравенству треугольника  $d - d' \leq r$ . Такие же неравенства справедливы для всех элементов, поэтому  $\mu - \mu' \leq \sigma - \sigma'$ . Учитывая выведенное неравенство  $\sigma \leq \mu$ , получаем, что  $\sigma' \leq \mu'$ , т. е.  $\sigma'$  — минимально  $\blacklozenge$ .

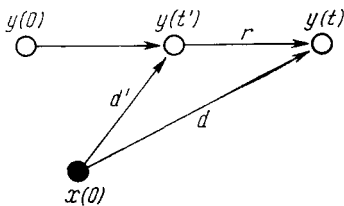


Рис. 3.8. Монотонная зависимость от частоты регистрации промежуточных положений элементов

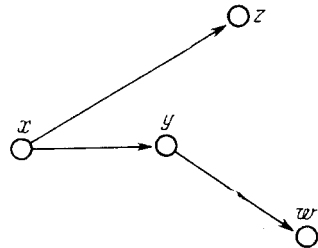


Рис. 3.9. Несогласованное сопоставление:  $x \rightarrow y$ ,  $x \rightarrow z$ , но  $y \rightarrow w$

Указанное свойство монотонности можно применить для вычисления соответствия, например, при удалении неверных вариантов сопоставления с помощью проверки промежуточных вариантов сопоставления на согласованность. Предположим, что элемент  $x$  на  $S1$  сопоставляется с элементом  $y$  на следующем кадре  $S2$  и с  $z$  на третьем кадре  $S3$ . Если соответствие установлено правильно, то эти варианты сопоставления должны быть согласованы, т. е.  $y \rightarrow z$ . На рис. 3.9 показано несогласованное сопоставление: в процессе установления соответствия  $x \rightarrow y$ ,  $x \rightarrow z$ , но  $y \rightarrow w$ . Если процесс установления соответствия будет допускать только согласованные между собой варианты сопоставления, то этим он может уменьшить число неверных сопоставлений. Такого рода проверку на согласованность можно проводить для любой схемы установления соответствия независимо от наличия у нее свойства монотонности. Однако из наличия этого свойства вытекает, что ситуации типа «ложной тревоги», когда переход  $x \rightarrow z$  правилен, но отвергается, являются крайне маловероятными. Наблюдения над процессом установления соответствия, используемого зрительной

системой человека, указывают на то, что эта система не пользуется такими проверками на согласованность. Но похоже, что в соответствии со свойством монотонности работа системы, отвечающей у человека за установление соответствия, по-видимому, монотонно улучшается с повышением частоты регистрации промежуточных положений элементов.

**Вид функции  $q(v)$  и некоторые следствия.** *Предпочтение, отдаваемое ближайшему соседу.* Уже не раз отмечалось, что всегда, когда это можно сделать, не нарушая других условий, процесс установления соответствия в зрительной системе человека стремится сопоставить каждый элемент с его ближайшим соседом. На рис. 3.10а, к примеру, элемент  $Z$  может перейти либо в  $Y1$ , либо в  $Y2$ . Оба варианта сопоставления являются допустимыми, по-

скольку в обоих случаях элемент  $Z$  переходит к единственному выходному элементу ( $X, V, Z$ ), а в каждом из них — к единственному входному элементу ( $Y1, Y2$ ). В подобной ситуации предпочтение всегда будет отдавать сопоставлению элемента  $Z$  с его ближайшим соседом. Иначе говоря, если на рис. 3.10а  $d_1 < d_2$ , то вариант сопоставления  $Z \rightarrow Y1$  предпочтительнее варианта  $Z \rightarrow Y2$ . На рис. 3.10б  $Z$  будет сопоставляться с  $Y2$ , поскольку его сопоставление с ближайшим соседом  $Y1$  приведет к возникновению недопустимого варианта сопоставления.

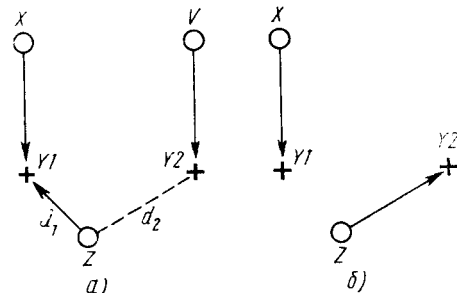


Рис. 3.10. Предпочтение, отдаваемое ближайшему соседу: на рис. б) в противоположность а)  $Z$  не связывается в пару со своим ближайшим соседом, поскольку результирующее сопоставление не будет являться минимальным покрытием

Может показаться, что предпочтение, оказываемое ближайшему соседу, указывает на то, что в процессе установления соответствия используется допущение о том, что низкие скорости более вероятны, чем высокие. В отношении низких скоростей это выглядит сомнительно: хорошо, пусть очень высокие скорости маловероятны, но ведь нет никаких оснований считать, что скорость, скажем, 1 град/с встречается значительно чаще (или же реже), чем скорость, скажем, 0,5 град/с. Но тогда чем же можно объяснить то несомненное предпочтение, которое при всех скоростях отдается ближайшему соседу элемента?

В рамках теории минимального отображения ответ состоит в том, что ближайший сосед должен получать весовое предпочтение почти независимо от того распределения вероятностей скоростей, которое характеризует окружающие его элементы. Напомним, что функция  $q(v)$ , которая должна быть минимизирована в процессе установления соответствия, была определена как



— $\log p(v)$ , где  $p(v)$  — распределение вероятностей скоростей в плоскости изображения. Обозначим через  $\rho(v)$  распределение вероятностей пространственных скоростей элементов. Иными словами, если наугад выбрать движущийся элемент, то вероятность того, что его скорость (в пространстве) заключена между  $a$  и  $b$ , составляет  $\int_a^b \rho(u) du$ . В предположении пространственной изотропности (т. е. того факта, что элементы с равной вероятностью могут двигаться в любом направлении) это распределение вероятностей скоростей в плоскости изображения будет пропорционально величине

$$\int_v^{\infty} \rho(u) (u^2 - v^2)^{-1/2} du. \quad (3.23)$$

(Это выражение справедливо для ортогональной проекции. При переходе к центральной проекции требуется незначительное изменение<sup>3,12</sup>.) Предпочтение, отдаваемое ближайшим соседям функцией  $p(v)$ \*, значительно весомей, чем предпочтение, отдаваемое функцией  $\rho(v)$ . Если функция  $\rho(v)$  монотонно убывает, то и  $p(v)$  тоже монотонно убывает. Однако  $p(v)$  может отдавать предпочтение своим ближайшим соседям даже в том случае, когда, как это имеет место, например, в диапазоне низких скоростей, сравнительно более высокие скорости являются (в трехмерном пространстве) более вероятными, чем сравнительно более низкие. Это объясняется наличием связи между пространственными скоростями и их проекциями. Если на проекции скорость равна  $v$ , то истинная пространственная скорость  $u$  должна быть, по крайней мере, равна  $v$ , но может оказаться и любой большей по значению скоростью. Следовательно, чем меньше проекция скорости, тем шире множество пространственных скоростей, которые могли бы

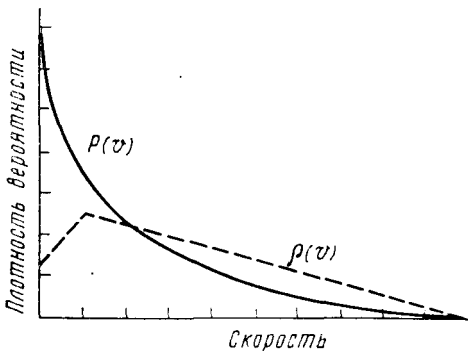


Рис. 3.11. Распределения вероятностей проекционной скорости  $p(v)$  и скорости в плоскости изображения  $\rho(v)$

дать эту проекцию. На рис. 3.11 приведен пример, в котором выбранная функция  $\rho(v)$  вначале возрастает, а затем убывает. С другой стороны, функция  $p(v)$ , подсчитанная по формуле (3.23), при всех скоростях оказывает заметное предпочтение ближайшим соседям.

Если  $\rho(v)$  монотонно убывает, то напрашивается очевидная стратегия установления соответствия, состоящая в сопоставлении

\*  $p(v)$  вычисляется по формуле (3.23). — Прим. ред.

элементов по принципу «ближайшего соседа», когда каждый элемент на одном из двух данных кадров переходит в своего ближайшего соседа на другом кадре. В самом деле, при достаточно малых скоростях, когда сопоставление по принципу «ближайшего соседа» приводит к взаимно-однозначным отображениям, минимальное отображение совпадает с сопоставлением по принципу «ближайшего соседа». При более высоких скоростях сопоставление по этому принципу уже не оптимально, так как оно часто отклоняется от взаимно-однозначности и приводит к многочисленным расщеплениям и слияниям. Минимальное отображение является методом сопоставления по принципу «ближайшего соседа», усовершенствованным в двух отношениях. Во-первых, вместо того чтобы минимизировать сами длины пути, оно минимизирует функцию стоимости этих путей. Как мы уже видели, их значения совпадают только в области малых скоростей, а при более высоких скоростях длины путей уже не являются истинной мерой правдоподобия данного варианта сопоставления. Во-вторых, минимальное отображение объединяет минимизацию и требуемую степень предпочтительности взаимно-однозначных отображений. Таким образом, минимальное отображение можно рассматривать как такое усовершенствование метода сопоставления «по ближайшему соседу», которое расширяет диапазон скоростей, в котором это отображение можно применять. Мы уже отмечали тот факт, что сходимость минимального отображения имеет глобальный характер, но скорость сходимости зависит от выбора начальной точки. Для ускорения сходимости в качестве возможной начальной точки может быть взят вариант сопоставления по принципу «ближайшего соседа»; тогда последующие итерации будут улучшать этот вариант, подводя его к минимальному отображению.

*Конфигурация Тернуса и S-образная форма функции  $q(v)$ .* Конфигурация Тернуса для кажущегося движения была введена в разд. 2.4. Она состоит из пары точек  $A$  и  $B$ , предъявляемых в быстрой последовательности со второй парой точек  $B$  и  $C$ . Точки  $A$ ,  $B$ ,  $C$  лежат на одной горизонтальной прямой. При различных условиях соответствие между ними может восприниматься одним из двух способов. При способе установления соответствия по связанному элементу пара точек  $(A, B)$  передвигается вправо как единое целое (т. е. воспринимается соответствие  $A \rightarrow B, B \rightarrow C$ ). При способе по соседнему элементу  $B \rightarrow B$ , а  $A$  часто совершает перескок» и сопоставляется с  $C$ .

Если при малых скоростях функция  $q(v)$  выпукла\*, то минимальное отображение в диапазоне малых скоростей должно приводить к осуществлению способа установления соответствия по связанному элементу. Тогда оба пути, проходимые элементами при сопоставлении, равны  $d$ . При способе по соседнему элементу один путь равен 0, другой —  $2d$ . Из выпуклости  $q(v)$  вытекает, что

\* То есть  $q''(v) > 0$ . — Прим. пер.

$$q(0) + q(2d) > 2q(d). \quad (3.24)$$

Следовательно, способ по связанному элементу минимизирует сумму  $\Sigma q_i$ .

Как будет видно, есть основания считать, что в области высоких скоростей функция  $q(v)$  вогнута. Таким образом, функция  $q(v)$ , как показано на рис. 3.12, приобретает S-образную форму. Это означает, что если  $v$  достаточно велико, то метод минимального отображения будет отдавать предпочтение способу установления соответствия по соседнему элементу. На рис. 3.12 с помощью тонкой прямой линии указана точка перехода от одного способа к другому, в которой  $q(0) + q(2d) = 2q(d)$ . Аналогичным образом из S-образной формы функции вытекает правильное предсказание: пересекающиеся траектории, у которых длина одной

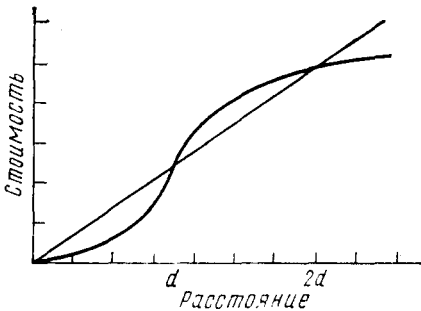


Рис. 3.12. S-образная форма функции  $q(v)$  может привести к смене способа установления соответствия в конфигурации Тернуса

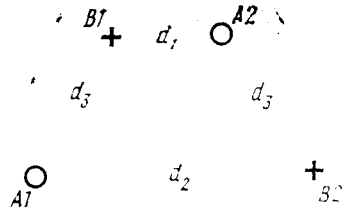


Рис. 3.13. Бистабильная конфигурация. Суммарный путь одинаков у обоих способов установления соответствия:  $d_1 + d_2 = 2d_3$

больше длины другой, по мере роста скорости возникают со все большей легкостью. Другие аспекты поведения бистабильных экспериментальных изображений также соответствуют приведенному выше анализу для конфигурации Тернуса. Рассмотрим, к примеру, конфигурацию на рис. 3.13, в которой  $d_1 + d_2$  в 2 раза больше, чем  $d_3$ . Это изображение бистабильно и обнаруживает два способа установления соответствия. При малом МСИ преобладают сопоставления  $A1 \rightarrow B2$ ,  $A2 \rightarrow B1$ , а при большом — сопоставления  $A1 \rightarrow B1$ ,  $A2 \rightarrow B2$ . В противоположность конфигурации Тернуса в данном случае эти способы не могут быть разделены на движения элементов и движения групп [Пэнгл и Пиччиано, 1976]. Однако с точки зрения минимального отображения эта конфигурация похожа на конфигурацию Тернуса. Если при обоих способах суммарный путь одинаков, то из S-образной формы функция  $q(v)$  вытекает, что в области выпуклости предпочтительнее тот способ сопоставления, при котором оба элемента проходят одинаковые расстояния, а в области вогнутости — способ с не-одинаковыми расстояниями. Заметим еще, что если на втором

редъявляемом кадре оба элемента  $B$  сместить влево на расстояние  $h$ , то для способа по связанному элементу суммарный путь уменьшается (на величину  $2h$ ), а для способа по соседнему элементу суммарный путь остается неизменным. Как предсказывает метод минимального отображения, при таком варианте конфигурации Тернуса с ростом  $h$  способ установления соответствия по связанному элементу становится все более предпочтительным.

### 3.7. Экспериментальное определение функции $q(v)$

Оптимальный вариант сопоставления друг с другом двух наборов точек определяется функцией  $q(v)$ . (В  $q(v)$  может входить константа  $k$  видоизмененного минимального отображения.) Если зрительная система применяет метод установления соответствия, похожий на минимальное отображение, то есть ли возможность выявить ту функцию  $q(v)$ , которую использует зрительная система? Иными словами, можно ли, изучая сопоставление, проводимое зрительной системой, единственным образом определить  $q(v)$ ? Предположим, что существует функция  $q'(v)$ , которая всегда дает те же варианты сопоставления, что и  $q(v)$  (т. е. сумма  $\Sigma q'_i$  минимальна всякий раз, когда минимальна  $\Sigma q_i$ ). Такая функция  $q'(v)$  не должна отличаться от  $q(v)$ . Можно, однако, показать, что  $q(v)$  в принципе определима с точностью до единственного масштабного множителя и что если ограничиться рассмотрением только взаимно-однозначных отображений, то  $q'(v)$  неотличима от  $q(v)$  тогда и только тогда, когда для некоторых констант  $a$  и  $b$  выполняется условие  $q'(v) = aq(v) + b$ . Это означает, что, рассматривая только взаимно-однозначные отображения, мы можем определить  $q(v)$  с точностью до коэффициентов линейного преобразования\*. Приводимая далее процедура дает пример того, как можно определить функцию  $q(v)$ .

Будем использовать бистабильные экспериментальные изображения, подобные конфигурации Тернуса. Если бистабильная конфигурация обладает двумя вариантами сопоставления  $m$  и  $m'$  примерно одинаковой вероятности, то  $\Sigma q_i \approx \Sigma q'_i$ , где  $\Sigma q_i$  — общая стоимость  $m$ , а  $\Sigma q'_i$  — общая стоимость  $m'$ . При смене способов восприятия конфигурации Тернуса имеем

$$\hat{q}(0) + \hat{q}(2v_1) = \hat{q}(v_1) + \hat{q}(v_1). \quad (3.25)$$

Положим  $\hat{q}(0) = 0$ , а  $\hat{q}(v_1) = 1$ . Тогда  $\hat{q}(2v_1) = 2$ . Здесь применено обозначение  $\hat{q}(v)$ , а не  $q(v)$  для того, чтобы указать различие между функцией  $\hat{q}(v)$ , которая определяется из бистабильных конфигураций при условии  $\hat{q}(0) = 0$ ,  $\hat{q}(v_1) = 1$ , и истинной искомой функцией  $q(v)$ . Теперь для нахождения новых значений  $\hat{q}(v)$  можно использовать  $v_1$  и  $v_2 = 2v_1$ . На рис. 3.13, к примеру, можно избирательно менять  $d_3$ , оставляя  $d_1$  и  $d_2$  постоянными, до тех пор,

\* В оригинале: «с точностью до линейной функции», что не совсем точно.—  
Ом. пер.

пока не получим бистабильную конфигурацию, т. е. пока варианты сопоставлений  $A1 \rightarrow B1$ ,  $A2 \rightarrow B2$  и  $A1 \rightarrow B2$ ,  $A2 \rightarrow B1$  не станут равновероятны. (Для противодействия эффектам группирования может потребоваться введение фоновых элементов.) Для этой конфигурации будем иметь  $\hat{q}(v_1) + \hat{q}(v_2) = \hat{q}(v_3) + \hat{q}(v_3)$ . Следовательно,  $\hat{q}(v_3)$  также определено<sup>3,13</sup>. С теоретической точки зрения этот метод может быть расширен для того, чтобы определить  $\hat{q}(v)$  на всюду плотном множестве значений  $v$  (так как между любыми двумя данными значениями всегда имеется еще одно значение). Следовательно, функция  $\hat{q}(v)$  может быть измерена. Вернемся теперь к нашей исходной функции  $q(v)$ , которая должна быть линейной функцией от  $\hat{q}(v)$ , т. е.  $q = a\hat{q} + b$ . Для определения аддитивной константы можно воспользоваться такими бистабильными конфигурациями, у которых два возможных варианта сопоставления отличаются друг от друга числом элементов, переходящих друг в друга при сопоставлении. Пример такой конфигурации представлен на рис. 3.14. По мере постепенного увеличения расстояния  $h$  при постоянстве других расстояний распределение соответствий будет переходить от вида, представленного на рис.

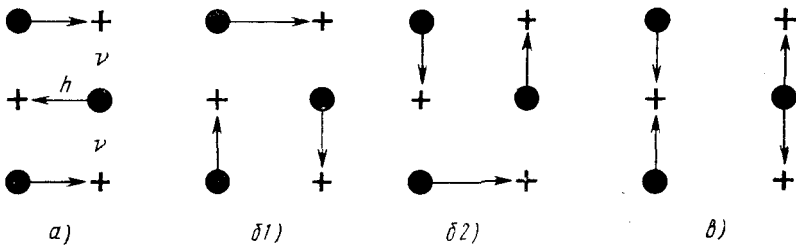


Рис. 3.14. По мере увеличения расстояния  $h$  преобладающий вид распределения соответствий изменяется от  $a$ ) к  $б$ ) и окончательно к  $в$ ) (на изображениях  $б$ ) и  $в$ ) общее число связей неодинаково)

3.14а, через вид, представленный на рис. 3.14б, к виду, представленному на рис. 3.14в. Заметим, что при  $h = 2v$  распределение соответствий на рис. 3.14б и на рис. 3.14в имеет одинаковую суммарную длину. Если функция стоимости не содержит штрафного множителя, то и при низких скоростях (большой МСИ и большое время предъявления) суммарная стоимость должна быть той же самой (см. разд. 3.6). Это должно бы привести к мультистабильной конфигурации, в которой конфигурации  $б1$ ,  $б2$  и  $в$  были бы равновероятны. В действительности же преобладает распределение соответствий, представленное на рис. 3.14б, и для получения мультистабильности  $h$  должно быть больше, чем  $2v$ . Из этого вытекает, что штрафной множитель, по всей вероятности, входит в функцию стоимости и этим «настраивает» процедуру установления соответствия на вариант сопоставления, имеющий наименьшее число связей.

Если распределения соответствия на рис. 3.14б и  $v$  равновероятны, то

$$q(v) + q(v) + q(v) + q(v) = q(v) + q(v) + q(h). \quad (3.26)$$

Подставив  $q = a\hat{q} + b$ , получим  $2\hat{q}(v_2) + b/a = \hat{q}(h)$ . Функции  $\hat{q}(v)$  и  $\hat{q}(h)$  уже известны, известно и  $b/a$ . Поскольку  $q$  можно определить только с точностью до масштабного множителя, то приходим к выводу, что  $q = c(\hat{q} + b/a)$ , где  $c$  — произвольная постоянная.

Для определения  $q(v)$  могут быть применены и другие экспериментальные процедуры. В принципе  $q(v)$  можно определить (с точностью до масштабного множителя) на основе тех вариантов сопоставления, которые осуществляются зрительной системой. Проведенные автором некоторые предварительные исследования функции  $q(v)$  хорошо вписываются в систему представлений, очерченную выше. Впрочем, для более точного определения зависимости стоимости от скорости и выяснения ее зависимости от других параметров потребуется более совершенная система опытов, что и обсуждается в следующем разделе.

### 3.8. Обобщения

До сих пор наше обсуждение было сосредоточено на установлении соответствия между двумя кадрами, каждый из которых состоял из точек одинаковой яркости. В данном разделе мы так расширим концепцию поиска наиболее вероятного варианта сопоставления между элементами с помощью простого локального процесса, чтобы она охватывала разные типы элементов и движение непрерывного характера.

#### 3.8.1. Расширение множества элементов

Как описывалось в гл. 1 и 2, в множество опорных элементов, сопоставляющихся друг с другом в процессе установления соответствия, входят такие образования, как фрагменты краев, отрезки прямых и округлые пятна. Главное новшество, которое вводится при расширении множества опорных элементов, заключается в том, что оптимальное отображение уже не определяется одними лишь расстояниями и временными интервалами. На правдоподобие варианта сопоставления между двумя элементами в общем случае влияют и другие параметры, такие как ориентация, длина и контраст. Эти параметры влияют на степень правдоподобия сопоставления двух данных элементов, и, следовательно, они участвуют в процессе установления соответствия через свою функцию стоимости  $q$ . В зрительной системе человека воздействие на  $q$  может оказываться также и местоположением элементов в зрительном поле, о чем свидетельствуют периферийные эффекты прироста и выравнивания, описанные в гл. 2. Однако минимальное отображение по-прежнему минимизирует сумму  $\sum (q_{ij} + k)x_{ij}$  и может быть найдено с помощью простого локаль-

ного вычисления, обсуждавшегося в разд. 3.4. Эквивалентным образом если сродство является мерой правдоподобия различных вариантов сопоставления (как это было объяснено в разд. 3.4), то оптимальное отображение в общем случае получается при максимизации суммарного сродства.

Точка зрения, что: 1) вариант сопоставления, который выбирается процессом установления соответствия в зрительной системе человека, действительно может быть предсказан на основе максимизации суммарного сродства и 2) относительная степень влияния различных параметров сродства согласуется с соображениями правдоподобия, — подтверждается некоторыми экспериментальными данными.

**Пример 1.** Если элементы на диагонали одинаковы, а элементы, расположенные вдоль боковых сторон, разные, то вероятность возникновения пересекающихся траекторий (см. разд. 3.6) возрастает. Аналогично, управляя степенью сходства входящих в изображение элементов (изменяя их ориентацию, длину и интенсивность), можно по выбору отдавать предпочтение способу установления соответствия либо по соседнему элементу, либо по связанному элементу [Фернбергер, 1934]. Другое интересное свойство конфигурации Тернуса, которое согласуется с рассуждениями, основанными на соображениях минимальной стоимости (максимального сродства), — это влияние межкадрового сходства. Уменьшение этого сходства усиливает способ установления соответствия по связанному элементу. Пусть первый кадр состоит из двух вертикальных полос, а второй — из двух горизонтальных, тогда способ по связанному элементу будет сильнее, чем в стандартной конфигурации Тернуса, в которой все элементы одинаковы. Пэнгл и Пиччиано (1976) обнаружили, что разница в контрасте (темные точки на первом кадре, светлые — на втором) дает похожий вклад в способ по связанному элементу и что этот вклад может объясняться уменьшением степени сходства (хотя в данном случае может сказаться и явление группирования). На рис. 3.15 схематически проиллюстрировано влияние дополнительно внесенного различия между кадрами на функцию стоимости  $q$ . В стандартной конфигурации Тернуса суммарная стоимость способа установления соответствия по соседнему элементу составляет  $q(0) + q(2v)$  (если считать, что  $q$  — функция только от  $v$ ). Суммарная стоимость способа по связанному элементу составляет  $q(v) + q(v)$  (рис. 3.15a). В случае несходства между кадрами

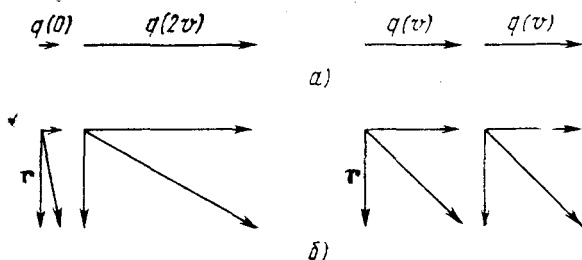


Рис. 3.15. Различие между кадрами усиливает способ установления соответствия по связанному элементу; вектор  $\mathbf{r}$  представляет дополнительную стоимость, возникающую благодаря различию

Элементы второго кадра отстоят от элементов первого не только в пространстве, но и в другом измерении, например по ориентации или контрасту. Из рис. 1.156 следует, что дополнительное разнесение элементов можно рассматривать как прибавление к исходным расстояниям постоянного вектора  $g$ , принадлежащего другому измерению. (Обоснование этого факта не зависит от взаимной перпендикулярности осей, имеющей место на рисунке.) Нетрудно убедиться в том, что в результате внесения дополнительного различия получается  $q_1 + q_2 > q_3 + q_4$ , и, следовательно, теперь общая стоимость минимизируется с помощью способа установления соответствия по связанному элементу.

**Пример 2.** Из табл. 2.1 следует, что степень сроства между двумя короткими отрезками зависит от пространственного углового расстояния между ними. Если считать, что пространство изотропно, и если функция стоимости на самом деле является мерой относительной вероятности вариантов сопоставления, то именно такой зависимости и следовало ожидать.

### 3.8.2. Непрерывное движение

Цель этого раздела — распространить наш анализ с дискретного предъявления двух кадров на случай непрерывного движения. Будет видно, что и в непрерывном случае оптимальное решение может быть найдено с помощью простой вычислительной сети, которая является простым расширением сети, описанной в разд. 3.4, а в случае дискретного предъявления она сводится к этой сети<sup>3.14</sup>.

В непрерывном случае время изменяется непрерывно, но можно считать, что местоположение элементов меняется не непрерывно, т. е. элементы могут находиться в дискретных ячейках изображения. Здесь в противоположность дискретному случаю возникновение и исчезновение элементов в разных ячейках уже не синхронизированы. Рассмотрим случай, когда  $n$  элементов движутся в промежутке от  $t=t_0$  до  $t=T$ . В качестве первого шага к рассмотрению общего случая предположим, что в моменты  $t=t_0$  и  $t=T$  на изображении находятся все  $n$  элементов.

В этом случае допустимые варианты сопоставления будут следующими. Каждый из  $n$  элементов в момент  $t_0$  обладает одной связью, соединяющей его с последующим элементом (т. е. с элементом, который появится в более поздний момент). В момент  $t=T$  каждый из  $n$  элементов обладает одной связью, соединяющей его с предыдущим элементом. В промежуточные моменты каждый элемент обладает двумя связями: одна — с предыдущим, другая — с последующим элементами. В соответствии с гипотезой независимости оптимальный вариант сопоставления — это тот вариант, который минимизирует  $\sum q_i$  на множестве всех допустимых вариантов сопоставления, где  $i$  пробегает все связи в данном варианте сопоставления.

В ситуации с двумя кадрами установление соответствия было эквивалентно задаче отыскания покрытия для двудольного графа, причем при определении локального строения вычислительного процесса двудольность играла центральную роль. Теперь опреде-



лим установление соответствия в непрерывном случае как покрытие двудольного графа и выясним локальное строение вычислительной процедуры непрерывного установления соответствия. Будем рассматривать каждый элемент как пару, состоящую из истока и стока. Истоки ответственны за установление связей с следующими элементами, а стоки — с предыдущими. В качестве своих первоначальных кандидатов каждый исток имеет стоки всех последующих элементов в некоторой пространственной окрестности. Тогда граф всевозможных пар становится двудольным, причем одну долю образует множество всех истоков, а другую — множество всех стоков. Как и раньше, оптимальный вариант сопоставления можно найти путем решения задачи:

$$\begin{aligned} &\text{Минимизировать } \sum q_{ij}x_{ij} \text{ при условиях: } \sum x_{ij} \geq 1 \text{ для } & (3.27) \\ &\text{всех } i, \text{ пробегающих множество истоков,} \\ &\text{и } \sum x_{ij} \geq 1 \text{ для всех } j, \text{ пробегающих множество стоков.} \end{aligned}$$

Задача, сформулированная таким образом, снова эквивалентна задаче оптимального установления соответствия при условии, что  $x_{ij} = 1$ , если  $i$ -й исток сопоставляется с  $j$ -м стоком, и  $x_{ij} = 0$  в противном случае. Поскольку на двудольном графе  $x_{ij}$  заведомо может принимать только значения 0 и 1, то эти две формулировки эквивалентны. Заменяв  $q_{ij}$  на  $q_{ij} + k$ , как это делалось раньше, можно также настроить оптимальное сопоставление на минимальное число соединений. Тогда эта задача с формальной точки зрения будет тождественна задаче из разд. 3.4. Следовательно, локальный процесс [см. уравнения (3.8) и (3.9)] будет сходиться к оптимальному варианту сопоставления.

Для непрерывного варианта установления соответствия при движении функция стоимости  $q$  зависит помимо ранее упомянутых параметров и от разнесенности элементов *во времени*. Можно ожидать, что в зрительной системе  $q_{ij}$  (стоимость связи между элементами  $i$  и  $j$ ) возрастает по мере увеличения разделяющего элемента временного интервала. При прочих равных параметрах (включая скорость) предпочтение будет отдано тому варианту сопоставления, который минимизирует временной промежуток. Например, вариант сопоставления двух элементов, разнесенных на расстояние  $d$  и временной интервал  $t$ , будет сильнее варианта сопоставления двух элементов, разнесенных на расстояние  $2d$  и временной интервал  $2t$ . По мере роста временного интервала, разделяющего два элемента, убывает правдоподобие варианта сопоставления этих элементов, связанное со стоимостью  $q$  обратно пропорциональной зависимостью. Если этот временной интервал превышает некоторый верхний предел  $\tau$ , то эти два элемента уже больше не рассматриваются в качестве кандидатов на сопоставление. Вместо общего для всех элементов временного интервала в пределах которого устанавливается соответствие (в предыдущем примере это был интервал от  $t_0$  до  $T$ ), каждый элемент выбирает себе потенциальные парные элементы в пределах времен-

бо интервала  $t$ . Для такой сети нет ни первого, ни последнего изображения; оптимальное соответствие вычисляется непрерывно, по мере поступления входных элементов.

### **3.9. О связях между разными подходами к установлению соответствия, развитыми в главах 2 и 3**

Как в гл. 2, так и в гл. 3 исследуется процесс установления соответствия, но для этого применяются разные подходы. Представляется интересным рассмотреть связи между этими главами тем, чтобы с некоторой общей точки зрения сравнить представленные в них различные подходы.

В гл. 2 упор делается на непосредственное изучение процесса установления соответствия у человека и интерпретацию наблюдаемых его проявлений. Такого рода исследование в конечном счете сталкивается с двумя непростыми проблемами. Первая связана с невозможностью объяснить процесс на основе полного знания соответствий между входом и выходом. Для случая сложных процессов такое непосредственное изучение может оказаться практически неприемлемым, а может быть, и невозможным. Например, для процесса, состоящего из последовательности операций (как программа для ЭВМ), наиболее эффективным будет описание, составленное на языке этих элементарных операций. Аналогичным образом для того чтобы сделать удобным в обращении описание процесса установления соответствия, надо, вероятно, выявить основные его компоненты и их взаимодействия. Но каким образом можно выяснить внутреннюю структуру сложного процесса? (Если процесс линеен, то для него имеются методы исследования «черных ящиков», но для классов нелинейных систем разработано только несколько ограниченных по своим возможностям методов.) Если этот процесс можно рассматривать как процесс решения некоторой вычислительной задачи, то приемлемый подход заключается в попытке проанализировать вычислительную задачу. Это исследование состоит из уточнения цели вычисления последующим изучением возможных методов достижения этой цели. Предположим, к примеру, что перед исследователем поставлена задача: выяснить внутреннее устройство электронного калькулятора. Если исследователь знает, что это приспособление используется для выполнения арифметических операций, то для определения направления исследования он будет располагать важным ориентиром. Без него же исследование может оказаться весьма нелегким.

Вторая проблема, связанная с принятым в гл. 2 подходом, относится к выяснению того, что же именно нас интересует. Как отмечалось во введении, теория того, что подлежит вычислению, является неотъемлемой частью изучения систем переработки информации. Что касается процесса установления соответствия, то можно, к примеру, задаться вопросом не только о виде конкурирующих взаимодействий, описанных в гл. 2, но также о том, ка-

кую функцию они выполняют. Таким образом, изучение с вычислительной точки зрения направлено на то, чтобы как дополнять, так и направлять экспериментальный подход. Конечно, само изучение в свою очередь тоже направляется экспериментальными данными, подобно тому как материал, изложенный в гл. 3, был подготовлен и направлялся результатами гл. 2.

В заключение хотелось бы кратко изложить общий взгляд на схему установления соответствия, вытекающий из гл. 2 и 3. Входное изображение обрабатывается так, что элементарные опорные символы образуются раньше, чем устанавливается соответствие. Тогда для процесса установления соответствия изменяющееся изображение может быть представлено как поток поступающих на вход процесса опорных символов. Между поступающими опорными символами непрерывно устанавливается минимальное отображение, которое приводит к наиболее правдоподобному варианту сопоставления. Основой этого варианта сопоставления является функция сродства (см. гл. 2) или в двойственной формулировке функция стоимости (см. гл. 3). Функция стоимости отражает некие «статистические свойства универсального множества элементов»<sup>3,15</sup>; эта функция измеряет относительную вероятность различных вариантов перехода элементов друг в друга, и, кроме того, благодаря ей сопоставление в определенной мере «настроено» на взаимно-однозначные отображения. При этом назначение локальных взаимодействий состоит в минимизации суммарной стоимости (с учетом штрафа за отклонения от взаимно-однозначности) и тем самым в оптимизации варианта сопоставления.

## ЧАСТЬ II

# ТРЕХМЕРНАЯ ИНТЕРПРЕТАЦИЯ

### КРАТКОЕ СОДЕРЖАНИЕ

После того как изображения одного и того же физического объекта с разных точек наблюдения сопоставлены описанным выше процессом нахождения соответствия, следующим шагом является интерпретация двумерных изменений внешнего вида объекта. Эта интерпретация включает в себя вычленение объектов в меняющейся сцене и восстановление их трехмерной структуры, движения и изменения. В гл. 4 и 5 определяются и исследуются ответственно два различных процесса интерпретации, а именно: *определение структуры по движению и определение движения по структуре.*

**Глава 4.** Зрительная система человека обладает замечательной способностью правильно интерпретировать меняющиеся проекции движущихся неузнаваемых объектов. Эта глава начинается с исследования проблемы интерпретации с вычислительной точки зрения, т. е. с того, как такая задача *может быть* решена. Полученные результаты затем используются для построения схемы интерпретации структуры по меняющемуся изображению движущихся объектов.

В разд. 4.1 формулируется проблема интерпретации структуры по движению. В нем описывается демонстрационный эксперимент с *двумя цилиндрами*, иллюстрирующий возможность зрительной системы человека выводить трехмерную структуру из меняющегося изображения даже в том случае, когда каждая статическая проекция сама по себе не содержит трехмерной информации.

В разд. 4.2 критикуются три предложенных ранее модели процесса интерпретации. Это — оценка глубины по скорости, связь кинетической глубины с некоторыми детерминантами и непосредственное восприятие Гибсона.

В разд. 4.3 утверждается, что для облегчения задачи интерпретации требуется дополнительная информация, и вводится понятие *рефлексивных ограничений.*

В разд. 4.4 в качестве рефлексивного ограничения для интерпретации структуры по движению формулируется *предположение сохранения формы.* Анализ этого предположения основан главным образом на теореме о структуре по движению, в которой утверждается, что структура, построенная из четырех не лежащих

в одной плоскости точек, однозначно восстанавливается по трем ортогональным проекциям. Полученные результаты используются для разработки схемы, способной реально осуществлять интерпретацию, которая выполняется без предварительных знаний или же распознавания объектов в рассматриваемой сцене.

В разд. 4.5 предложенный метод распространяется на случай перспективной проекции. Рассматриваются два обобщения: прямой перспективный метод, который учитывает иную геометрию (см. подразд. 4.5.1), и полярно-параллельный метод, в котором в случае перспективы используется параллельная схема для получения локальных аппроксимаций (см. подразд. 4.5.2). Теоретические рассуждения, так же как и экспериментальные данные, наводят на мысль, что в зрительной системе человека процесс интерпретации основан на принципах, подобных имеющим место в полярно-параллельной схеме.

В разд. 4.6 обсуждаются соответствующие экспериментальные данные с целью проверки применимости схемы, основанной на сохранении формы и ее пригодности в качестве психологической теории восприятия структуры по движению.

**Глава 5.** Здесь обсуждается второй процесс интерпретации, а именно схема определения движения по структуре. Этот процесс участвует в интерпретации пространственного движения, но, в отличие от схемы определения структуры по движению, в этом случае требуется, чтобы трехмерную структуру сцены можно было воспринимать по статическим проекциям.

В разд. 5.1 приводятся данные, свидетельствующие в пользу существования процесса интерпретации движения по структуре, и исследуются некоторые его свойства. В отличие от схемы определения структуры по движению, рассмотренной в предыдущей главе, процесс определения движения по структуре основывается на статическом трехмерном описании наблюдаемой сцены. Используя трехмерные «метки», приписанные элементам последовательных изображений, строится пространственное движение элементов, которое зрительная система «заполняет» до непрерывного.

В разд. 5.2 рассматриваются различные явления, связанные с иллюзией «окно Эймса». Этот раздел включает вторую часть книги, поскольку в нем собраны вместе различные аспекты описанных двух схем интерпретации.

## ГЛАВА 4

### ИНТЕРПРЕТАЦИЯ СТРУКТУРЫ ПО ДВИЖЕНИЮ

#### 4.1. Восстановление структуры по параллельным проекциям

Следующий этап после установления соответствия связан с интерпретацией воспринимаемых двумерных изменений в терминах трехмерной структуры и характера движения наблюдаемых объектов.

В рамках крайнего эмпирического подхода к проблеме интерпретации можно было бы для интерпретации попытаться использовать имеющийся опыт, связанный с наблюдаемым объектом или с похожими объектами. Так, при восприятии вращающегося кубика различные его изображения связываются на основе предшествующего опыта с одним и тем же кубиком, рассматриваемым под разными углами зрения. Однако о недостаточности такого рода эмпирического объяснения свидетельствует тот факт, что изображение движущихся объектов можно правильно интерпретировать даже в случае, когда объекты незнакомы и каждое статическое изображение сцены не содержит информации об их трехмерной структуре. Эта удивительная способность зрительной системы человека интерпретировать структуру и движение неузнаваемых объектов была предметом многочисленных психологических исследований<sup>4,1</sup>. Впервые ее систематическим изучением занялись Уоллок и О'Коннэлл (1953), которые исследовали, по их терминологии, «кинетический эффект глубины». В этих экспериментах незнакомый объект вращался за матовым экраном, а на другой стороне экрана наблюдалась его тень, являющаяся результатом параллельной проекции объекта на экран. (Термин «параллельная проекция» будет использоваться в этой главе наравне с ортогональной проекцией и обозначать проекцию, которая образуется с помощью параллельных лучей, перпендикулярных плоскости изображения.) В большинстве случаев наблюдатели могли дать правильное описание структуры и движения закрытого экраном объекта даже в том случае, когда на каждом статическом изображении объект не распознавался.

В первых экспериментах по исследованию кинетического эффекта глубины главным образом использовались объекты, сделанные из проволоки, проекции которых представляли собой связанную совокупность отрезков линий. В более поздних исследованиях [например, Уоллок и О'Коннэлл, 1953; Уайт Мюзер, 1960; Грин, 1961; Браунштейн, 1962; Йоханссон, 1974, 1975] было установлено, что по предъявляемым изображениям, содержащим движущиеся несвязанные элементы, также можно воспринимать трехмерную структуру. (Термин «элемент» будет использоваться для обозначения любых идентифицируемых характерных точек, таких как изолированные точки, концы отрезков линий или текстурные элементы.) Как правило, использовалось небольшое число элементов (один [Хофстен, 1974], два [Борьессон и Хофстен, 1972], три [Борьессон и Хофстен, 1973], до шести [Браунштейн, 1962]) или элементы, находящиеся на плоских поверхностях [Гибсон и Гибсон, 1957; Фиандт и Гибсон, 1952; Гибсон и другие, 1959; Гибсон, 1965]. В следующем разделе описывается демонстрационный эксперимент, обобщающий в какой-то степени указанные выше случаи на изображения с большим числом точек, расположенных в виде двух неплоских конфигураций. Этот пример проиллюстрирует восприятие структуры по движению и поможет сформулировать вычислительные задачи, лежащие в основе такого восприятия.

**Демонстрационный эксперимент с двумя цилиндрами.** На экране дисплея, управляемого ЭВМ, показывалась ортогональная проекция двух коаксиальных цилиндров. Каждый цилиндр задавался с помощью 100 точек, лежащих на его поверхности. Общая ось цилиндров была направлена вертикально (рис. 4.1). Трехмерные координаты точек хранились в памяти ЭВМ, а их ортогональная проекция на фронтальную плоскость вычислялась и показывалась на экране.

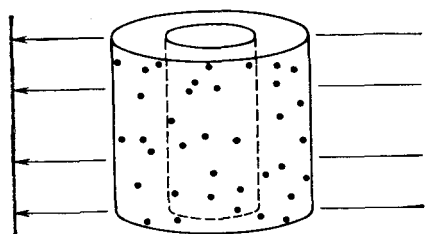


Рис. 4.1. Проекция двух цилиндров (вид сбоку). Границы цилиндров на предъявляемом изображении не показывались

Воображаемые цилиндры затем поворачивались (угол поворота не превышал на каждом шаге  $10^\circ$ ), после чего вычислялась и показывалась на экране их новая проекция (время предъявления и межстимульного интервала составляли соответственно 100 и 40 мс). На проекции плотность точек возрастала по краям цилиндров, но вследствие того что на изображении были представлены два цилиндра, изменения плотности были сложными и бесполезными для выделения трехмерной структуры предъявляемых объектов. Каждое отдельное статическое изображение, таким образом, выглядело почти случайным набором точек. Однако, когда наблюдалась меняющаяся проекция, движущиеся поперек экрана элементы воспринимались как два вращающихся цилиндра, форма и углы поворота которых легко определялись. Как разбиение сцены на объекты, так и соответствующая трехмерная интерпретация были основаны в этом случае исключительно на движении, поскольку каждое отдельное изображение не содержало информации ни для такого разбиения, ни для восстановления структуры. Каждая предъявленная картинка представляла собой неизвестное и неузнаваемое изображение двух цилиндров, что позволяет сделать вывод о том, что знание объекта и распознавание его на изображении не являются необходимыми условиями интерпретации движения.

Важно обратить внимание на два ограничения в рассмотренном эксперименте. Во-первых, положение оси вращения в пространстве оставалось при движении постоянным. Однако и в похожих экспериментах, в которых ориентация оси вращения менялась скачками (на  $30^\circ$  и более) после каждой картинке, все же воспринималась трехмерная структура (ср.: «кувыркающееся движение» [Грин, 1961]). Во-вторых, использовались дискретные стимулы, приводящие к кажущемуся движению. Похоже, что это не является существенным для процесса интерпретации. Трехмерную структуру можно воспринимать как по непрерывному, так и по кажущемуся движению, и последующий анализ применим к обоим случаям.

Итак, проблема интерпретации структуры по движению включает в себя разбиение меняющегося изображения рассматриваемой сцены на изображения объектов и восстановление их пространственной структуры и пространственного движения, даже когда трехмерная информация в одиночном изображении не представлена явно.

Термин «пространственное движение» означает здесь преобразование (вращение с последующим сдвигом) одного известного положения данного тела в другое. Термин «структура», примененный выше, следует отличать от понятия *абсолютной* глубины. Речь идет об определении лишь только относительной *внутренней* глубины. Реальные размер и расстояние могут оставаться неопределенными, но все внутренние расстояния в объекте определяются точно до единого масштабного коэффициента. (Иными словами, не существует способа различать монокулярные проекции большого, близко расположенного объекта и объекта, находящегося далеко, по размеру пропорционально большего; см. также приложение 2.)

Цель этой главы состоит в рассмотрении вычислительных задач, лежащих в основе интерпретации структуры по движению, в построении схемы, способной реально проводить такую интерпретацию. Например, если задана проекция двух цилиндров, такая схема разобьет сцену на два объекта и восстановит их правильную трехмерную структуру, а также их движение в пространстве. Прежде чем предложить решение, коротко обсудим три предложенные ранее модели, которые выглядят неадекватными. Критические замечания помогут акцентировать внимание на некоторых особенностях изучения зрительного восприятия вообще.

#### 4.2. Критика существующих подходов

Один из подходов к проблеме интерпретации структуры по движению состоял в оценке действительной глубины\* отдельных точек на основе их скоростей: чем выше скорость элемента, тем ближе он находится [Гельмгольц, 1910; Браунштейн, 1962; Гершгер и Старжец, 1974]. В соответствии с этими работами восстановление структуры по движению аналогично стереоскопическому восприятию глубины. Основная идея заключается в том, чтобы заменять последовательные кадры (восприятие движения) одномоментными (стереоскопическое восприятие), допуская, что процесс нахождения соответствия играет роль бинокулярного влияния, а значения смещений играют роль несоответствий. За исключением некоторых особых случаев (например, чистого сдвига наблюдателя относительно стационарной окружающей среды), этот подход не может быть корректным, поскольку в общем случае по смещениям или скоростям нельзя определить глубину. В сцене могут находиться объекты, движущиеся в различных направлениях и с произвольными скоростями. При этом может отсутствовать связь между скоростью (или смещением) и глубиной. Рассмотрим демонстрацию ленты конвейера, которая позволяет ценить воздействие различий в скорости на воспринимаемую глубину. В этом случае точки и короткие отрезки линий перемещаются по экрану дисплея через три области *a*, *b* и *c*, как это

\* То есть расстояние до глаз наблюдателя. — Прим. ред.



представлено на рис. 4.2а, причем границы между областями на экране не показываются. В области  $b$  скорость равна  $v$ , а в областях  $a$  и  $c$  она равна  $v'$ , где  $v' < v$ . На дисплее имитируется фронтальный вид ленты конвейера, которая сверху выглядит так, как на рис. 4.2б. Что должно восприниматься в этом случае?

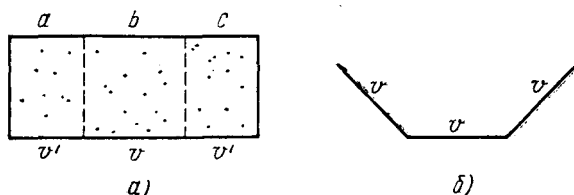


Рис. 4.2. Демонстрация ленты конвейера: фронтальное изображение элементов, движущихся с постоянной скоростью вдоль ленты конвейера (а), вид сверху (б)

Теория смещений предсказывает, что  $a$  и  $c$  будут казаться лежащими в одной плоскости, а  $b$  — в другой. Другая возможность состоит в том, что лента конвейера будет восприниматься так, что скорость элементов вдоль их траекторий останется неизменной<sup>4,2</sup>. Ни одно из указанных предположений не подтверждается опытом. Движение элементов в фронтальной плоскости воспринимается следующим образом: при входе в  $b$  они ускоряются, а при выходе из  $b$  замедляются<sup>4,3</sup>. Хотя в примере с двумя цилиндрами нельзя определить глубину, зная скорость, все же в каждом цилиндре скорости элементов меняются в соответствии с глубиной. Поэтому, вообще говоря, можно было бы предположить, что *группирование* элементов (в тела) должно предшествовать анализу глубины. Вероятно, это является одной из причин того, почему эффект группирования при движении изучался как самостоятельная проблема. В гештальтпсихологии имеется, например, понятие «группирование по схожести свойств», которое включает в себя группирование на основе близости скоростей. Поттер (1974) использовал аналогичный критерий: элементы  $i$  и  $j$  с соответствующими скоростями  $v_i$  и  $v_j$  объединяются, если  $|v_i - v_j|$  не превышает некоторый предварительно заданный порог. Гибсон [1965] изучал проблему группирования и убедился, что близость скоростей не является адекватным критерием объединения. Впоследствии он занимался поиском класса двумерных преобразований, обеспечивающих группирование. Пример с двумя цилиндрами иллюстрирует сложности, связанные с процессом группирования по движению. В каждом цилиндре имеются точки с разнообразными скоростями из некоторого интервала. Вместе с тем точки с абсолютно одинаковыми скоростями относятся к разным объектам.

Другая модель интерпретации структуры по движению была предложена впервые в работе, посвященной кинетическому эффекту глубины [Уоллок и О'Коннэлл, 1953], а также и в более поздних работах [Уоллок и другие, 1956; Янссон и Йоханссон,

73]. Способность к восприятию структуры по движению объяслась в рамках «эффекта», вызванного меняющимися одновременно по длине и ориентации линиями и контурами. Если бы в изображении учитывались только реально существующие линии, объяснение оказалось бы явно неверным, так как описанная выше структура из несвязанных точек может быть восстановлена их движению. Поэтому пришлось учитывать и воображаемые линии, связывающие идентифицируемые точки [Уоллок и О'Коннэлл, 1953]. Но и скорректированного предположения, т. е. того, что при восприятии трехмерной структуры используются линии, виртуальные линии, а также контуры, у которых меняются длина и ориентация, конечно, недостаточно. Рассмотрим, например, случайное движение несвязанных элементов на плоскости. Длина и ориентация виртуальных линий, соединяющих их, постоянно меняются, однако не наблюдается никакой ясной трехмерной структуры. Приведенное выше предположение оказывается необходимым лишь в тривиальном смысле: условия Уоллока и О'Коннэлла нарушаются только такими двумерными изменениями изображения, как преобразования, сохраняющие форму (изображений, но не самих трехмерных объектов), и однородное масштабирование. Если же структура трехмерного объекта не восстанавливается по единственной проекции, то не удивительно, что однородное смещение, вращение или масштабирование самого изображения оказывается недостаточным для определения неизвестной структуры.

Ранее было высказано предположение [Уоллок и О'Коннэлл, 1953; Борьессон и Хофстен, 1972], что для преобразований, в которых все элементы в изображении движутся параллельно, не возникает кинетический эффект глубины. Однако при демонстрации двух цилиндров все элементы движутся параллельно горизонтальной оси и трехмерная структура воспринимается.

С кинетической моделью глубины прямо связано понятие *непосредственного восприятия* Гибсона. В соответствии с теорией Гибсона удастся правильно интерпретировать меняющиеся проекции: «... не из-за того, что мы образовали связи между оптически и элементами, и даже не из-за того, что мозг объединил эти оптические элементы, но потому, что сетчатка глаза чувствительна к изменениям как таковым» [Гибсон, 1957; с. 294].

В более общем смысле, с точки зрения Гибсона, восприятие состоит в непосредственном связывании стимулов и перцептов\*. Таким образом, в этой теории нет места и необходимости в процессах, осуществляющих вывод, интерпретацию или вычисление [Гибсон, 1959]. Использование подхода непосредственного восприятия к зрению приводит к теории, в основных аспектах прямо противоположной предлагаемому нами вычислительному подходу.

\* Перцепт — образ, создаваемый объектом с помощью зрительного аппарата. Ниже будет видно, что это несколько укрупненная по сравнению с той реакцией на стимул единица восприятия. — *Прим. ред.*

Изучение применимости подхода непосредственного восприятия к проблеме интерпретации структуры по движению обнаруживает два основных его недостатка. Первый относится к понятию *непосредственности*. Второй касается разделения процесса восприятия объекта на отдельные этапы образования пар стимулов и перцептов. Обсудим эти недостатки более подробно.

*Непосредственное восприятие.* Термин «непосредственный» является относительным в смысле его зависимости от конкретной системы. Если рассматривается некоторая система, то любой внешний сигнал, который ее достигает, можно считать непосредственно воспринимаемым. Сигналы о тепле или касании, выработанные периферическими рецепторами, психолог мог бы рассматривать как непосредственно воспринимаемые. В то же время психолог, изучающий, например, внутренний механизм корпускулы Майснера (тактильный рецептор), не может считать непосредственной связь между раздражением и реакцией рецептора. Поэтому оценка отношения между зрительными стимулами и восприятием структуры по движению как непосредственного не означает отсутствия в зрительном восприятии некоторого промежуточного процесса. Получается, что при изучении перцептивной системы восприятие структуры по движению скорее следует считать посторонним фактором. Такая точка зрения встречает два очевидных возражения. Во-первых, трехмерная интерпретация нуждается в объяснении, даже если для некоторого другого процесса или системы ее можно считать непосредственной. Во-вторых, сомнительным выглядит определение границ перцептивной системы способом, при котором результат процесса интерпретации считается непосредственно заданным.

*Разделение восприятия на пару стимул — перцепт.* В рамках вычислительного подхода восприятие структуры по движению объяснялось бы как процесс получения трехмерного описания по двумерным изменениям изображения. В рамках же подхода непосредственного восприятия последнее объясняется с помощью эффектов, вызываемых некоторыми простыми детерминантами. Например, в случае кинетической модели глубины этими детерминантами были бы линии изображения с меняющимися длиной и ориентацией. Подход непосредственного восприятия ведет к разделению процесса восприятия объекта на составляющие его эффекты, поскольку *истинную* трехмерную структуру объекта нельзя представить в виде единственного эффекта, связанного, например, с линиями, длина и ориентация которых меняются. Для интерпретации структуры по движению были введены такие перцепты, как трехмерность, сохранение формы и связность, которые были связаны с различными детерминантами. Такая декомпозиция приводит к перцептивно-бихевиористической теории зрения, в которой бихевиористические пары стимул — реакция заменяются парами стимул — перцепт. Такая теория объясняет зрительное восприятие с помощью составления некоторой таблицы, в которой связываются друг с другом стимулы и перцепты. По каждому

стимулу или классу стимулов можно найти перцепты, которые тот стимул в состоянии вызвать с учетом приписанных им потенциальных весов, характеризующих степень эффективности этого стимула для вызывания каждого перцепта. Табличный метод выглядит явно неподходящим для определения структуры по движению. Восприятие такой структуры, как, например, два цилиндра, намного богаче, чем какая-либо совокупность эффектов. Иначе говоря, при наблюдении движущегося объекта усваивается нечто большее, чем эффекты трехмерности, сохранения формы и связности, а воспринимаются реальные формы и движение объекта. Поэтому правильное объяснение интерпретации структуры по движению, вероятно, должно осуществляться с помощью процесса восстановления истинной трехмерной структуры по двумерным изменениям, а не в терминах набора эффектов, связанных с характерными стимулами.

Для изучения процесса получения трехмерной структуры по двумерным преобразованиям изображения будем следовать общей схеме, изложенной во введении, и разработаем вычислительную теорию для восстановления структуры по движению. Следующая задача, которой нужно заняться, сводится к тому, как трехмерную структуру и движение объектов можно получить из двумерных изменений их проекций.

### 4.3. Рефлексивные ограничения

Главная задача, лежащая в основе интерпретации структуры по движению, состоит в неоднозначности такой интерпретации. Иначе говоря, единственной структуры и движения, соответствующих данному двумерному преобразованию, не существует. Например, при демонстрации двух цилиндров движущихся на экране элементы воспринимаются лежащими на вращающихся трехмерных цилиндрах, а не, как это есть на самом деле, подвергаемыми двумерному сдвигу. Далее, такие плоская и двухцилиндровая интерпретации не являются единственными, соответствующими предъявляемому преобразованию. Это только два из бесконечного числа пространственных движений элементов, которые могли бы дать ту же самую двумерную проекцию.

Для того чтобы избавиться от такой неопределенности, схема интерпретации должна содержать некоторые ограничения, отбрасывающие большинство из возможных трехмерных интерпретаций и приводящие к единственному решению, которое в большинстве реальных случаев соответствует действительности. Эти ограничения можно рассматривать и как набор *неявных предположений* о физическом мире, выполнение которых приводит к правильному решению. Наоборот, систему можно ввести в заблуждение и построить зрительные иллюзии, создав условия, при которых неявные предположения не выполняются.

Неявные предположения имеют место в различных задачах восприятия и поэтому заслуживают отдельного обсуждения. Их

необходимость вытекает из того факта, что отношение между объектами и их изображениями не является взаимно-однозначным.

Зрительный процесс можно рассматривать состоящим из двух взаимно-обратных отображений: зрительное кодирование и зрительное декодирование. Сначала структура видимого внешнего мира кодируется в оптическом сигнале, который достигает глаза. Затем этот сигнал декодируется зрительной системой для того, чтобы получить первоначальную структуру внешнего мира. Этап декодирования осложняется тем, что большое число отличающихся друг от друга структур может отображаться в один и тот же оптический сигнал. В условиях многозначности, присущей этому процессу, совершенное декодирование становится невозможным. Однако во многих случаях некоторые способы кодирования встречаются чаще, чем другие<sup>4.4</sup>. Этап декодирования может, опираясь на такое предпочтительное кодирование, использовать особый метод декодирования с помощью внутренних ограничений, которые отдают предпочтение наиболее вероятным интерпретациям.

Таким образом, внутренние ограничения рассматриваются как предположения, отражающие те свойства физического мира, которые управляют этапом кодирования. Именно в этом смысле я называю такие ограничения *рефлексивными*. Введенный термин поможет отличать рефлексивные ограничения, используемые зрительной системой от ограничений, наложенных на нее (см. введенное в гл. 3 ограничение на локальность).

Рефлексивная роль внутренних ограничений становится ясной при сравнении рефлексивных ограничений с *законом прегнантности* в гештальтпсихологии. Этот закон объясняет, каким образом многозначные стимулы приводят к определенным перцептам: *«Психологическая организация всегда будет настолько же «хорошей», насколько это позволяют преобладающие условия ... [где «хорошая»] ... охватывает такие свойства, как регулярность, симметрия, простота и др.»* [Коффка, 1935, с. 110].

Этот общий закон в дальнейшем разбивается на ряд принципов прегнантности: замыкания, хорошего продолжения, хорошей формы и простейшего пути. В соответствии с позицией гештальтпсихологии организационные принципы прегнантности аналогичны правилам, которым подчиняется механика. Механические системы достигают равновесия под действием механических сил. По аналогии в формулировке гештальтпсихологии «устойчивые пространственные организации сохраняются в результате взаимодействия сил» [Коффка, 1935, с. 439]. Эти силы представляют собой силы отталкивания и притяжения в головном мозге, а результирующие перцептивные структуры «уменьшают внешние силы организации» [там же, с. 171]. Считается, что такие организационные принципы не относятся к свойствам окружающей среды, а связаны с особенностями структуры мозга. Следовательно, такого рода рассуждения ведутся с точки зрения физических механизмов, а не реально существующих ограничений, налагаемых на осмысленные вычисления.

Принципы прегнантности иногда заменялись утверждениями, что система восприятия подчиняется определенным принципам минимума (например, [Эттив, 1972]). Можно было бы заметить, однако, что принципы минимума являются только одним из возможных формальных выражений существующих правил и что вопрос о том, какой из них предпочтительнее, не имеет значения с точки зрения научных результатов [Фейнман, 1965]. Например, механика Гамильтона, основанная на принципе наименьшего действия Мопертюи, является механикой Ньютона в формулировке, использующей принцип минимума. Основной вопрос поэтому заключается не в том, можно ли представить «организационные» правила восприятия в виде принципа минимума, а что собой представляют эти правила и какие ограничения они отражают.

К понятию рефлексивных ограничений ближе концепция «бесознательных заключений» Гельмгольца [Гельмгольц, 1910] и концепция образования перцептивной гипотезы Грегори [Грегори, 1970]. Как и рефлексивные ограничения, эти понятия также подразумевают применение ограничений или предположений для правильного декодирования неоднозначных стимулов. Новое состоит в том, что рефлексивные ограничения отражают универсальное использование знаний об общих свойствах объектов, а не избирательное использование специфических знаний о заданном объекте или ситуации. Как утверждалось в разд. 3.2, значение знаний общего, рефлексивного вида преобладает на ранних стадиях зрительного анализа<sup>4,5</sup>.

Основные требования к рефлексивным ограничениям заключаются в том, что они должны быть действительно *ограничивающими*. Иными словами, такие ограничения должны уменьшать число интерпретаций до (предпочтительно) одной и быть правдоподобными в том смысле, что почти всегда должны выполняться в окружающем мире. Возвращаясь к задаче интерпретации структуры по движению, необходимо, следовательно, предположить существование рефлексивных ограничений и показать, что эти ограничения удовлетворяют указанным требованиям и поэтому приводят к правильной интерпретации.

#### 4.4. Схема интерпретации структуры по движению

Первый этап анализа представлен на рис. 4.3. Ищется процесс интерпретации, который на основании получаемых двумерных проекций определяет трехмерную структуру. Поскольку эта задача недоопределена, полезно считать, что информация, на основе которой осуществляется интерпретация, поступает из двух источников: двумерных преобразований и рефлексивных ограничений.

Ограничение, предлагаемое для интерпретации структуры по движению, будет называться *предположением о сохранении формы* и заключается в следующем: *любая совокупность элементов, подвергаемых некоторому двумерному преобразованию, имеющему*

единственную интерпретацию в виде некоторого движущегося в пространстве твердого тела (сохраняющего свою форму), должны быть интерпретированы как такое движущееся тело.

Отдавая предпочтение интерпретациям, сохраняющим форму, следуем нескольким работам [например, Уоллок и О'Коннэлл, 1953; Гибсон и Гибсон, 1957; Грин, 1962; Хэй, 1966; Йоханссон, 1975] и используем похожий «принцип трехмерности» Йоханссона

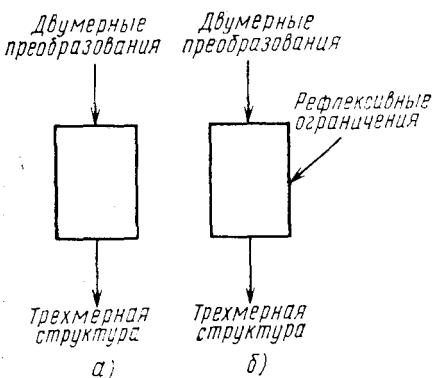


Рис. 4.3. Интерпретация структуры по движению использует два источника: двумерные преобразования и рефлексивные ограничения

[Йоханссон, 1964, 1975], который обнаружил, что сохранение формы, вероятно, играет особую роль в восприятии движения. Однако такое «предпочтение сохранению формы» является только отправным моментом для настоящего анализа. В предположении о сохранении формы содержится больше, чем просто отражение процесса восприятия человеком сохраняющих форму объектов. Это предположение учитывает свойства физического мира с тем, чтобы прийти к правильной интерпретации недоопределенных стимулов. Следующий шаг анализа поэтому состоит в том, чтобы показать, как предположение о сохранении формы можно включить в схему интерпретации, способную правильно выводить структуру из движения.

**Включение предположения о сохранении формы в схему интерпретации.** Для того чтобы использовать предположение о сохранении формы, схема интерпретации должна быть в состоянии определить, имеет ли данный набор движущихся элементов только одну интерпретацию, сохраняющую форму. Можно начать интерпретацию, подвергая наборы элементов сцены следующему тесту на сохранение формы: *имеет ли этот набор единственную интерпретацию в виде движущегося в пространстве твердого тела?* Если ответ отрицательный (вследствие того, что имеется более чем одна интерпретация, сохраняющая форму, или вследствие того, что их нет вовсе), тогда элементам не приписывается структура. Если ответ положительный, тогда единственная обнаруженная структура связывается с этими элементами.

Предположение о сохранении формы требует, чтобы используемый тест применялся к небольшим наборам элементов сцены. Это было бы необходимо, если бы, например, сцена состояла из нескольких различно движущихся объектов. Если бы все элементы сцены сразу были подвергнуты тесту на сохранение формы, то можно было бы получить отрицательный результат просто потому, что элементы принадлежат к более чем одному объекту. Следовательно, тест на сохранение формы необходимо применять к тому,

то в дальнейшем будет называться *ядром* элементов, а именно к минимальному числу элементов, по которым еще можно определить однозначно их структуру. Посмотрим, что собой представляет такое ядро.

С тестом на сохранение формы непосредственно связаны две проблемы. Первая заключается в том, можно ли тест реализовать в виде вычислительного процесса. Иначе говоря, можно ли указать процедуру выяснения существования единственной интерпретации, сохраняющей форму, для набора движущихся элементов и определения соответствующей структуры. Вторая заключается в том, приведет ли предложенная процедура к правильной интерпретации предъявленных проекций. Как будет показано ниже, ответ на оба вопроса положительный: тест является вычисляемым, и процедура приводит к правильной интерпретации сцены.

Рассмотрим сначала вторую проблему, исследовав возможные пути, которые могут привести к неправильной интерпретации. Одна из возможностей появления ошибки возникает в случае, когда тест на сохранение формы отвечает «да» вместо «нет», а другая, когда тест отвечает «нет» вместо «да»\*. Первая ошибка вызывает появление ложных целей — точек, движущихся на самом деле независимо друг от друга, вместе с тем дающих единственную интерпретацию в виде движущегося твердого тела. В этом случае рассматриваемая схема интерпретации приводит к ложной структуре, на которую она случайно натолкнулась. Вторая ошибка связана с появлением фантомных структур — точек, принадлежащих на самом деле проекции некоторого движущегося объекта, вместе с тем связывающихся также с проекцией некоторого другого движущегося объекта. В этом случае появляется дополнительная (фантомная) структура, поэтому решение не является единственным и, следовательно, движущимся элементам не приписывается никакая структура.

Высокая эффективность схемы интерпретации, основанной на сохранении формы, обусловлена тем, что вероятности неправильной интерпретации обоих типов пренебрежимо малы. Сначала рассмотрим случай фантомных структур и покажем, что при некоторых простых предположениях он невозможен. Иначе говоря, если задана ортогональная проекция некоторого движущегося объекта, то не существует другого объекта и другого сохраняющего форму движения, совместимых с заданной проекцией. Это утверждение следует из теоремы об объектах, сохраняющих свою форму, которая называется *теоремой о структуре по движению*: *если даны три различные ортогональные изображения четырех точек, не лежащих в одной плоскости и принадлежащих к сохраняющей форму конфигурации, то структура и движение, совместимые с данными тремя изображениями, определяются единственным образом.*

\* В отечественной литературе эти ошибки называются соответственной «ложной тревогой» и «пропуском сигнала». — *Прим. пер.*



Эта теорема была первоначально сформулирована и доказана Ульманом [1977а, приложение 1] для пяти точек. Фремлин [1977, устное сообщение] показал, что эти требования могут быть ослаблены до четырех точек. Доказательство, основанное на обеих указанных работах, приводится в приложении 1.

В теореме о структуре по движению изображения получаются с помощью ортогональных проекций. Как показано, например, в экспериментах по кинетической глубине и при демонстрации двух цилиндров, зрительная система человека способна выводить структуру из ортогональных проекций. Именно этот случай сейчас будет рассмотрен. В разд. 4.5 полученные результаты будут распространены также на случай перспективных проекций.

Сформулированная теорема имеет двойное применение для процесса интерпретации структуры по движению. Во-первых, она устанавливает, что трехмерную структуру можно определить всего лишь по трем изображениям четырех точек. Это и есть минимальное ядро, с которым может работать схема интерпретации. Во-вторых, тот факт, что структура определяется однозначно, свидетельствует о том, что фантомные структуры невозможны. Отсюда следует, что такой тип неправильной интерпретации исключается.

Вторым типом неправильной интерпретации, как уже говорилось, являются ложные цели. Можно показать, что в нашем трехмерном мире ложные цели встречаются очень редко. Мала вероятность того, что три изображения движущихся без сохранения формы четырех точек приведут к интерпретации, сохраняющей форму. Действительно, если положения точек задаются совершенно точно, то эта вероятность равна нулю. (Доводы в пользу этого утверждения носят статистический характер и приводятся в работе [Ульман, 1977; приложение 1].)

Итак, из двух перечисленных выше неправильных интерпретаций фантомные структуры невозможны, а ложные цели имеют равную нулю вероятность появления. Следовательно, предлагаемая схема интерпретации практически защищена от указанных ошибок. Следует заметить, что поскольку используются ортогональные проекции, то объект определяется с точностью только до отражения относительно фронтальной плоскости. Это — естественная неоднозначность, поскольку объект, повернутый на угол  $\alpha$ , и зеркально симметричный ему объект, повернутый на угол  $-\alpha$ , имеют те же самые ортогональные проекции. Точно так же абсолютное расстояние до объекта и его сдвиг в глубину нельзя определить по соответствующей ортогональной проекции. Таким образом, интерпретация в случае ортогональных проекций позволяет: а) разбивать сцену на объекты; б) определять трехмерную структуру этих объектов с точностью до зеркального отображения и в) определять движение этих объектов (соотношение между начальными и конечными положениями и ориентацией) с точностью до сдвига в глубину.

Формулировка теоремы о структуре по движению, основанная на трех различных изображениях, ни в коей мере не означает, что движение исходного изображения должно быть дискретным. Если непрерывное движение достаточно длительное для того, чтобы можно было наблюдать три различных изображения (степень различия зависит от точности интерпретирующей системы), то в нем содержится достаточно информации для единственной интерпретации. В теореме формулируется этот математический факт без уточнения, как выделять такую информацию.

*Краткое изложение основных принципов.* Изложим основные принципы, лежащие в основе схемы интерпретации структуры по движению, разбив задачу интерпретации на две. Первая состоит в том, что трехмерная структура и движение являются недоопределенными, если пользоваться только двумерными преобразованиями. Эта трудность была преодолена использованием предположения о сохранении формы в качестве внутреннего ограничения. Вторая задача (определения истинного движения) состоит в том, что двумерные преобразования в данной сцене могли быть вызваны наличием в ней нескольких объектов, осуществляющих различные трехмерные движения. Эта трудность была преодолена в результате использования для интерпретации движения ядра элементов, которое, вообще говоря, должно принадлежать одному объекту.

**Реализация схемы интерпретации и ее применение к большим наборам элементов.** Доказательство теоремы о структуре по движению является конструктивным; в нем предлагается метод для построения схемы, способной реально определить структуру по движению. Такая схема была реализована в лаборатории искусственного интеллекта Массачусетского технологического института. Вместе с тем важно не забывать о различии между общим методом и конкретным алгоритмом, основанным на этом методе (обсуждение этого различия см. в [Марр, 1977a]). Схему определения структуры по движению можно реализовать с помощью различных алгоритмов. Для целей нашего рассмотрения наибольший интерес представляет общий метод, и поэтому алгоритм детально описываться не будет. Некоторые замечания о конкретной реализации можно найти в приложении, однако ряд положений, подтвержденных указанной выше реализацией, имеет смысл здесь привести.

*Плоские объекты.* Теорема о структуре по движению гарантирует единственное решение для случая трех изображений четырех элементов, не лежащих в одной плоскости. Это не означает, конечно, что о некопланарности требуется знать до применения теста на сохранение формы. Заметим также, что условие некопланарности является достаточным, но не необходимым: четыре элемента, лежащие в одной плоскости, могут иметь, вообще говоря, много решений или только одно. Это зависит от начальной ориентации плоского объекта и его последующих поворотов в пространстве. Если имеется только одно решение, то алгоритм

определяет структуру. В противном случае устанавливается лишь факт, что элементы лежат в одной плоскости, но ее ориентация и поворот остаются неоднозначно-определенными. Похожая ситуация возникает, когда наблюдаемый объект состоит только из трех точек. Одни конфигурации, образованные тремя точками, определяются, а другие не определяются однозначно тремя изображениями. Алгоритм, основанный на сохранении формы, можно применять к трем изображениям трех точек, но в этом случае получение единственного решения не гарантируется. Однако если интерпретация окажется единственной, то она будет обнаружена алгоритмом.

*Зависимость между числом точек и точностью.* В рассматриваемом алгоритме имеется определенная возможность обмена между числом используемых точек (или изображений) и требованиями к точности входной информации и точности вычислений. Если поступающие данные задаются с высокой точностью и вычислительный процесс имеет точность того же порядка, то достаточно четырех элементов. Для менее точного устройства (такого, каким, возможно, является зрительная система человека) может потребоваться и большее число элементов (или изображений) для надежной и точной интерпретации.

*Применение схемы к большим наборам точек.* Вследствие того что реальные сцены содержат обычно большое число элементов, необходимо оценить вычислительную сложность процесса интерпретации. Вопрос заключается в том, удастся ли осуществить вычисления при увеличении числа элементов до сотни или тысячи. Ответ таков: в реальных сценах объем требуемых вычислений похоже растет лишь линейно или приблизительно линейно с ростом числа точек. Более того, большую часть этого процесса можно реализовать параллельно, так что время вычислений может оказаться в значительной степени не зависящим от числа точек.

Для того чтобы исследовать ситуацию с большим числом элементов, предположим, что в некотором изображении имеется  $N$  элементов и  $K$  объектов. Можно разделить совокупность из  $N$  элементов на  $N/4$  группы, в каждой из которых содержатся четыре соседних элемента, и провести интерпретацию для каждой группы. Вычисления для различных групп не зависят друг от друга, и их можно было бы выполнять параллельно. Все группы, являющиеся внутренними относительно изображения объекта, т. е. те, в которых все четыре элемента принадлежат тому же самому объекту, будут удовлетворять тесту на сохранение формы и для них будет выделена структура. Дальнейшие рассуждения основываются на ожидании того, что точки, составляющие данный объект, не будут распределены случайным образом по всей сцене. В случае реальных непрозрачных объектов можно ожидать, что каждый объект будет иметь, по крайней мере, одну внутреннюю группу. (Случай нескольких прозрачных объектов является более сложным, но не настолько, чтобы нельзя было осуществить интерпретацию.) Прежде всего необходимо для каждого из  $K$  объ-

ектов получить совокупность внутренних точек, для которых определяются структура и движение, и совокупность граничных точек, которые пока остаются неопределенными. Далее, для каждой из остающихся граничных точек определяется ее принадлежность к одному из  $K$  объектов. Последнее можно выполнить для всех таких точек параллельно.

*Пример с двумя цилиндрами и прозрачными объектами.* В случае прозрачных объектов вычисления усложняются, так как выбранные локально точки могут принадлежать разным объектам, расположенным один позади другого. Однако если число видимых объектов мало в каждой малой области, то рост сложности ограничен. Алгоритм определения структуры по движению был использован в примере с двумя цилиндрами, в котором оба объекта были видны и находились один внутри другого. В центральной области (где изображения двух цилиндров перекрывались) почти все группы из четырех выбранных случайно точек содержали элементы обоих цилиндров и поэтому не имели интерпретации, сохраняющей форму. Однако для  $1/8$  этих групп все же существовала интерпретация, сохраняющая форму ( $1/16$  для каждого цилиндра). Все группы из периферийной области принадлежали только большему цилиндру и имели интерпретацию, сохраняющую форму. Таким образом, после первого шага трехмерная структура приписывается к более чем  $1/8$  всех точек. Второй шаг завершает интерпретацию, проводя ее описанным выше способом для граничных точек.

Таким образом, справедлив вывод, что предложенную схему интерпретации можно реально применять к сценам, содержащим большое число движущихся элементов. Если заданы три изображения некоторой сцены, содержащей движущиеся объекты, то вычислительный процесс скорее всего приведет к правильной интерпретации сцены. Элементы будут правильно объединены в объекты, а трехмерная структура этих объектов, так же как их движение в пространстве, будет определена. Интерпретация осуществляется на основе только движения, при этом не требуются знания о наблюдаемых объектах или распознавании статических проекций.

#### **4.5. Интерпретация перспективных проекций**

В предыдущем разделе была описана схема интерпретации, объясняющая способность зрительной системы определять структуру по параллельным проекциям. Однако воспринимаемая глазом проекция окружающего мира является не параллельной, а перспективной (рис. 4.4).

В этом разделе исследуется интерпретация перспективных преобразований. С целью расширения разработанной схемы интерпретации структуры по движению, основанной на предположении о сохранении формы, на перспективные проекции обсуждаются прямой перспективный и полярно-параллельный методы, которые

затем сравниваются с восприятием зрительной системы. В первом случае интерпретация проводится непосредственно и точно, являясь вариантом параллельной схемы интерпретации, измененной так, чтобы можно было учесть новую геометрию. Иначе говоря, эта схема интерпретации проверяет двумерные преобразования на их совместимость с перспективной проекцией, двужущейся и сохраняющей форму трехмерной конфигурации. Во втором случае, в полярно-параллельной схеме, используется параллельный метод без изменений, приводящий к приближенным результатам.

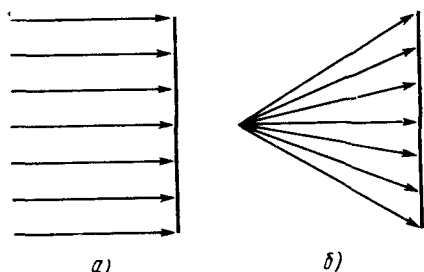


Рис. 4.4. Ортогональная (а) и перспективная (б) проекции на плоскость

Эти два метода сравниваются со зрительным восприятием человека. Можно предположить, основываясь на полученных результатах, что восприятие человеком структуры по перспективным проекциям ближе к полярно-параллельной, чем к прямой перспективной. Прямой перспективный метод превосходит это восприятие, но ценой большей восприимчивости к ошибкам и большей вычислительной сложности, тогда как

полярно-параллельный метод по возможности и ограничениям напоминает восприятие человека. В заключение полярно-параллельная схема интерпретации рассматривается как теория<sup>4,6</sup>, описывающая зрительное восприятие человеком структуры по движению.

#### 4.5.1. Прямая перспективная схема

В этом подразделе изучается непосредственное определение структуры и движения по перспективным проекциям. В общих чертах излагается прямая перспективная схема интерпретации, которая помогает установить следующие четыре факта:

1. Структуру и движение объектов можно определить по их перспективным проекциям. Для получения единственного решения в используемой схеме обычно достаточно трех изображений пяти элементов.

2. Свойственная перспективному методу восприимчивость к ошибкам связана с тем, что эффекты перспективы часто слабо выражены, и поэтому использующая их процедура зависит от точности входной информации и чувствительна к небольшим ошибкам. Это особенно остро проявляется в случае, когда интерпретацию приходится проводить локально, так как перспектива уменьшает размеры объектов. В случае ортогональных проекций было показано, что эффективная интерпретация, основанная на сохранении формы, должна выполняться локально на ядрах элементов для того, чтобы обойти элементы, удовлетворяющие тесту на сохранение формы, но принадлежащие разным объектам. Следова-

тельно, прямой перспективный метод не годится для интерпретации структуры по движению, основанной на сохранении формы.

3. Вычисления, необходимые для реализации прямой перспективной интерпретации, являются, вообще говоря, более сложными, чем для параллельного случая. Однако некоторые типы движения, в частности чистый сдвиг, удастся анализировать прямым перспективным методом так же успешно, как параллельной схемой.

4. В некоторых аспектах прямой перспективный метод интерпретации превосходит качественно зрительное восприятие. Видимо, это связано скорее не с используемым методом, а с тем, что способность к непосредственной интерпретации меняющихся перспективных проекций в зрительной системе человека отсутствует.

Настоящий раздел дополняется приложением 2, в котором кратко описывается алгоритм перспективной интерпретации. Суть алгоритма состоит в использовании *полярного уравнения*, на основе которого выполняется интерпретация движений, состоящих из поворотов относительно некоторой вертикальной оси с последующими произвольными сдвигами. Поэтому сначала будет рассмотрен именно такой тип движения. Результаты будут использованы как для более сложной ситуации — произвольного движения, так и для более простого случая — чистого сдвига.

**Поворот относительно вертикальной оси.** В приложении 2 описан алгоритм, способный определять структуру и пространственное движение элементов при выполнении следующих условий. Известны две перспективные проекции движущегося объекта, содержащего, по крайней мере, четыре точки. Два изображения различаются тем, что объект подвергают повороту относительно вертикальной оси и затем произвольному сдвигу. Сначала алгоритм находит угол поворота  $\alpha$ . Этот угол определяется из полярного уравнения

$$A \sin^2 \alpha + B \cos^2 \alpha + C \sin \alpha \cos \alpha + D \sin \alpha + E \cos \alpha = 0.$$

Хотя коэффициенты уравнения представляют собой довольно сложные выражения, они зависят только от проекций. Поэтому их можно вычислить по этим двум проекциям. После нахождения угла  $\alpha$  алгоритм может восстановить трехмерную структуру и движение с помощью формул, приведенных в приложении 2. В отличие от параллельного случая, полученное решение включает движение по направлению  $y$ , т. е. движение *в глубину*. Полярное уравнение дает практический и эффективный способ для проведения анализа в частном случае вращения относительно вертикальной оси (или относительно любой другой известной оси, лежащей в плоскости изображения) и последующего произвольного сдвига.

**Произвольное движение.** Алгоритм для случая вращения относительно оси, приведенный в приложении 2, был включен в общую схему, предназначенную для анализа произвольного движения. При этом приходилось вместо одного одновременно решать

большое число полярных уравнений. Требовался также значительный поиск, связанный с проверкой большого числа возможных осей вращения, поскольку угол поворота неизвестен. Поиск основан на предположении, что приближенно верные результаты получаются, если выбрана ось с ориентацией, достаточно близкой к истинной. Существуют также и другие методы, которые не требуют поиска и используют вместо этого итеративную процедуру решения системы нелинейных уравнений, связывающих вращение, сдвиг и трехмерные координаты элементов<sup>4,7</sup>. Во всяком случае, предложенный общий метод оказался практически неприемлем, но он и не был разработан для эффективного использования. Скорее он служил средством для излучения перспективного случая. Вместо того чтобы переносить теорему о структуре по движению на перспективные проекции, указанный алгоритм применялся для оценки минимальной информации, необходимой для проведения перспективной интерпретации. Обычно алгоритму было достаточно трех изображений пяти точек для определения единственно правильной структуры и движения. В отличие от ортогонального случая, часто было достаточно двух изображений пяти точек. (Для ортогонального случая недостаточно двух изображений любого числа точек.) В смысле минимальной информации, необходимой для надежной интерпретации, прямой перспективный алгоритм сравним, таким образом, с параллельной схемой и, как будет показано ниже, также с полярно-параллельным методом. Следует заметить, однако, что получаемые числа справедливы для конкретного алгоритма и зависят от его точности. Они не обязательно дают требуемый абсолютный минимум. С другой стороны, для менее точной системы (какой, возможно, является зрительная система человека) потребуются дополнительные изображения или точки для надежной интерпретации. Когда движение сводится к чистому сдвигу, применение прямой перспективной схемы намного облегчается и обычно требуется меньше информации. Рассмотрим этот случай.

**Чистый сдвиг.** Приведенное выше полярное уравнение можно использовать для построения простого теста, проверяющего, является ли движение наблюдаемого объекта чистым сдвигом. Подставив в полярное уравнение  $\alpha=0$  (отсутствие поворота), получаем  $B+E=0$ . Обратное справедливо в статическом смысле: если  $B+E=0$  (для четырех точек или более), то элементы, вероятно, принадлежат твердому телу, подвергнутому чистому сдвигу. Если чистый сдвиг обнаружен, то для вычисления структуры и движения можно использовать уравнения из приложения 2, положив  $\alpha=0$  (более подробно см. приложение 2). Таким образом, чистый сдвиг более всего подходит для схемы прямого перспективного анализа. Движение чистого сдвига поддается обнаружению, а сам сдвиг, так же как и трехмерная структура, определяется аналитическими выражениями.

**Сдвиг в глубину.** Частным случаем произвольного смещения является сдвиг по направлению к наблюдателю или от него, кото-

рый привлек особое внимание исследователей по восприятию движения [Гибсон и другие, 1959; Йоханссон, 1964; Мармолин, 1973 (1.11); Борьессон и Хофстен, 1972, 1973; Хофстен, 1974]. Такое особое внимание к этому случаю, вероятно, следует из сложившегося мнения, что поскольку определение информации о глубине играет важную роль в восприятии движения, то чистое движение в глубину в некотором смысле обостряет проблему интерпретации. Восприятие движения в глубину будет рассматриваться здесь в рамках предложенной перспективной схемы, использующей полярное уравнение. Полученные результаты будут сопоставлены с некоторыми другими теориями, а также с экспериментальными данными. Следует учесть два момента. Во-первых, настоящая схема интерпретации может успешно работать с удаляющимися и приближающимися объектами. Во-вторых, зрительная система человека отличается от рассматриваемой перспективной схемы интерпретации по способности выводить структуру по движению в глубину — факт, дающий серьезные основания для предположения, что зрительная система человека не обладает способностью к прямой интерпретации структуры по перспективным проекциям.

Для перспективной схемы интерпретации движение в глубину незначительно отличается от других сдвигов. Если в движении участвуют, по крайней мере, четыре точки, то алгоритм, основанный на полярном уравнении, правильно установит как их движение, так и структуру. Сначала выясняется, выполняется ли условие  $B+E=0$ , тем самым проверяется, участвуют ли рассматриваемые точки в чистом сдвиге. Затем вычисляются структура и движение. Следует обратить внимание на то, что проверка на чистый сдвиг проводится перед применением формул интерпретации. Для оценки этой предварительной проверки рассмотрим удаляющийся вертикальный стержень. Наблюдаемые изменения изображения заключаются в уменьшении вертикального размера, которое также можно связать с *поворотом* стержня в глубину (относительно горизонтальной оси). Этой ситуации внутренне свойственна неоднозначность, поэтому она не поддается единственной интерпретации. В работе о восприятии удаляющихся объектов [Мармолин, 1973] изучался упомянутый эффект укорочения стержня. (В реальных экспериментах в качестве стимулов использовались квадраты, но учитывались только одномерные изменения.) Движение в глубину вычислялось в предположении, что укорочение стержня вызывалось только его движением от наблюдателя. Эти результаты затем сравнивались с субъективными оценками испытуемых. Не удивительно, что хорошая корреляция отсутствовала. Как отмечал Мармолин, наблюдатели, встретившись с неоднозначными стимулами, приписывали укорочение стержня частично движению в глубину и частично повороту, показывая вследствие этого сильно различающиеся результаты. В противоположность описанной схеме перспективный метод не объяснял бы изменения внешнего вида объекта движения в глубину, если для



этого не было бы оснований. Даже если зрительная система человека и может определять структуру удаляющегося объекта по его движению, то в этом неоднозначном случае нельзя ожидать надежной интерпретации.

В более показательных экспериментах следовало бы оценивать восприятие структуры по движению в условиях, достаточных для того, чтобы гарантировать единственную интерпретацию. Были проведены такие эксперименты, и результаты свидетельствуют о том, что в определенных условиях люди не воспринимают структуру удаляющегося объекта, тогда как перспективная схема интерпретации работает в этих условиях с относительно малыми вычислительными затратами. В качестве одного из примеров предъявлялись изображения двух цилиндров (см. разд. 4.1, рис. 4.1), которые исследовались на восприятие движения, направленного от наблюдателя. Демонстрировались восемь изображений, которые имитировали (при рассматривании их с расстояния 80 см от экрана) постепенное удаление двух цилиндров диаметрами 25 и 50 мм, начальное и конечное расстояния до которых (до их общей оси) соответственно были равны 17 и 62,5 см. Изображения периодически показывались в прямом (удаление) и обратном (приближение) порядке, и эти движения сменялись циклически. Время предъявления каждого изображения составляло 70 мс, а межстимульный интервал был равен 30 мс. Хотя сжатие и расширение изображения привели к определенному впечатлению о движении в глубину, *структуру* двух цилиндров нельзя было определить. В противоположность такой неудаче перспективная схема интерпретации, примененная к этим проекциям, смогла определить трехмерную структуру двух цилиндров, а также их смещение в глубину.

Как теоретические рассуждения, так и экспериментальные наблюдения позволяют считать прямой перспективный метод мало пригодным для интерпретации перспективных проекций. Но если зрительная система восприятия движения не располагает специальными методами интерпретации перспективных проекций, то каким же образом воспринимаются реальные сцены, в которых проекции являются перспективными? Ответ заключается в том, что параллельная схема интерпретации структуры по движению, предложенная в разд. 4.4, может оказаться подходящей для решения этой задачи.

#### 4.5.2. Полярно-параллельная схема

Параллельную схему интерпретации структуры по движению можно почти без изменений использовать также для получения приближенных результатов в случае перспективной проекции. Локальный характер вычислений, как правило, приводит к разумным приближениям. Сначала будет дано объяснение того, почему такая аппроксимация возможна, а затем показано, как описанные локальные операции можно объединить в некоторую схему.

обладающую значительно бóльшими возможностями, чем отдельные ее компоненты.

Когда объект находится достаточно далеко, то его перспективные проекции можно считать слабо искаженными ортогональными проекциями. Поэтому вычислительную схему, используемую в случае ортогональных проекций, можно также применить для получения приближенных результатов в случае перспективных проекций. Термин «далеко» означает, что разности между расстояниями от наблюдателя до рассматриваемых точек несравнимо меньше самих расстояний. Пусть  $y_i$  и  $y_j$  обозначают расстояния от наблюдателя до соответствующих точек, тогда эти точки считаются «далекими», если  $|y_i - y_j|$  гораздо меньше  $y_i$  или  $y_j$ . Такое условие может выполняться независимо от действительного расстояния между наблюдателем и объектом, если удовлетворяются два требования: 1) объект имеет плотно распределенные видимые точки и 2) объект является непрерывным, т. е. для двух близлежащих точек на изображении соответствующие им точки объекта также расположены близко друг от друга. В этом случае можно использовать то преимущество локальных по своей сути вычислений, которым характеризуется параллельная схема, требующая для интерпретации только четыре точки. Если рассматриваемый объект удовлетворяет указанным двум требованиям, то, сужая в каждый данный момент поле зрения до четырех близлежащих точек и тем самым определяя локальную структуру, можно осуществить интерпретацию «далекого» объекта. Применение ортогонального алгоритма восстановления структуры по движению к перспективным проекциям показало его слабую чувствительность к искажениям и возможность практического использования для интерпретации перспективных проекций (более подробные количественные данные приведены в [Ульман, 1977а]).

Локальное применение ортогональной схемы интерпретации структуры по движению и последующее объединение локальных результатов отличаются от анализа всего изображения как одной ортогональной проекции. В случае ортогональной проекции существует ось проекции, общая для всех точек объекта. Применяя ортогональную схему к малым окрестностям и при этом рассматривая их как отдельные ортогональные проекции, можно считать, что ось проекции *меняет свое направление при переходе от окрестности к окрестности*. Для определения структуры ортогональным методом объект сначала разбивается на области, содержащие примерно четыре элемента. Затем анализируется каждая область, как если бы она была получена в результате применения ортогональных проекций, ось которых направлена на глаза к центру рассматриваемой области. Такой метод анализа перспективных проекций будет называться *полярно-параллельным*.

Объединение локальных результатов в глобальную интерпретацию приводит к схеме, которая превосходит свои компоненты в двух аспектах. Во-первых, пропадает неоднозначность в определении истинной и симметричной ей трехмерной структуры. Во-

вторых, появляется возможность интерпретации движения, направленного от наблюдателя. В случае ортогональной интерпретации трехмерная структура определяется с точностью до отражения относительно плоскости изображения, а поворот — с точностью до знака. Упомянутые две неоднозначности присущи этому методу и связаны друг с другом, так как повернутая на некоторый угол трехмерная структура имеет ту же самую ортогональную проекцию, что и зеркально симметричная ей трехмерная структура, повернутая на тот же по величине, но противоположный по направлению угол. Полярно-параллельная схема не оставляет места такой неоднозначности. Она может определить единственным образом как структуру, так и поворот. Ниже описывается, как осуществить однозначное определение структуры и способ, с помощью которого однозначно определяется направление поворота в полярно-параллельной схеме, сравнивается с предлагаемыми ранее объяснениями того же явления, в частности с помощью двигательного параллакса. Наконец, показывается, что при благоприятных условиях выполнима интерпретация движения в глубину, которая невозможна, если пользоваться чисто параллельной схемой интерпретации.

**Однозначное определение структуры.** В случае ортогональной проекции результат интерпретации определяется с точностью до отражения относительно фронтальной плоскости. Полярно-параллельная схема не приводит к такой неоднозначности. Это связано с тем, что, хотя структура определяется локально с точностью до отражения, требование глобальной согласованности делает возможным различение истинной и локально отраженных структур. Данное утверждение иллюстрирует рис. 4.5. На рис. 4.5а представлено ортогональное проектирование объекта  $O$ . Изображены две возможные интерпретации: истинная  $I$  (сплошная линия) и отраженная  $I'$  (пунктирная линия). На рис. 4.5б представлено перспективное проектирование того же объекта. Полярно-параллельная схема локально восстанавливает структуру в областях  $O_1$  и  $O_2$ . Каждая локальная операция интерпретирует структуру с точностью до отражения в двух локальных областях:  $I_1$  и  $I'_1$  для  $O_1$ , а  $I_2$  и  $I'_2$  для  $O_2$ . Поскольку направление проекции меняется от  $O_1$  к  $O_2$ , меняются также и направления отраженных структур. В то же время как  $I_1$  и  $I_2$  являются истинными интерпретациями локальных областей  $O_1$  и  $O_2$  и параллельны друг другу,  $I'_1$  и  $I'_2$  не удовлетворяют этому отношению.

Однако рис. 4.5, демонстрируя различие между ортогональной и полярно-параллельной интерпретациями, не объясняет, каким образом в этом случае отбрасывается неправильная структура. Имеющиеся у наблюдателя знания допускают, что истинной могла бы быть любая из двух интерпретаций, изображенных на рис. 4.5б. Правильную интерпретацию можно выделить, используя требование согласованности осей вращения поперек объекта. Для оси вращения, которая определяется двумя последовательными изображениями, как и для структуры в случае ортогональной проек-

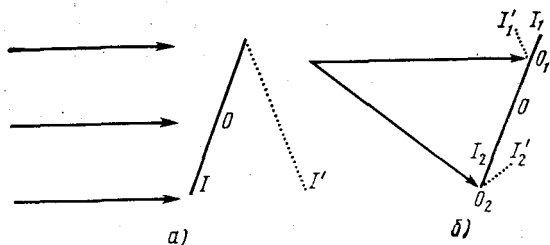


Рис. 4.5

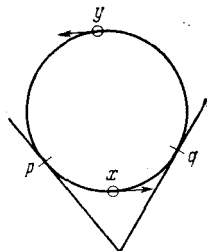


Рис. 4.6

Рис. 4.5. Неоднозначность интерпретации — глобальная в случае ортогональной (а) и локальная в случае перспективной (б) проекции. Во втором случае ее можно устранить

Рис. 4.6. Концы вращающегося стержня в условиях перспективной проекции (вид сверху), точки  $p$  и  $q$  являются неподвижными

ции, возможны два решения: истинное и отраженное. В то время как истинная ориентация остается постоянной для всего объекта, ориентация отраженной оси меняется при переходе от одной области объекта к другой. Требование наличия единственной оси вращения при повороте твердого тела, собственно, и приводит к единственной интерпретации. Как только однозначно определяется ось вращения, структура объекта восстанавливается также однозначно. Нахождение истинной оси облегчается при увеличении размера объекта. Причина заключается в том, что при восстановлении оси вращения по двум различным областям объекта истинные области будут иметь общую ось, в то время как расстояние между отраженными областями будет возрастать по мере увеличения углового расстояния между ними. Следовательно, максимальное расстояние между двумя ложными осями достигается в случае, когда соответствующие области объекта предельно удалены друг от друга. Таким образом, чем больше угловой размер объекта, тем легче и надежнее проводится распознавание истинной и ложной осей. Другой причиной, облегчающей выбор единственной структуры, является текстура объекта. Чем меньше расстояние между различаемыми элементами объекта, тем точнее восстановление истинной оси и тем легче она отделяется от ложной.

**Однозначное определение направления вращения с помощью движательного параллакса.** Направление вращения нельзя однозначно определить по параллельной проекции движущегося объекта. Как упоминалось выше, причина этого заключается в том, что проекция остается той же, если истинный объект заменить зеркально симметричным ему и повернутым в противоположном направлении. Однако такая инвариантность относительно отражения не выполняется для перспективной проекции. Рассмотрим ситуацию, когда два элемента  $x$  и  $y$  (например, два конца стержня) вращаются относительно своей средней точки, как это изображено на рис. 4.6. Точки  $p$  и  $q$ , лежащие на касательных, про-

веденных от наблюдателя к траектории движения, представляют собой две *неподвижные точки*: движение элементов в точках направлено вдоль их линий прямой видимости. Следовательно, в этих точках проекция скорости равна нулю. В положении, показанном на рис. 4.6, где  $y$  расположен «позади»  $x$ , угловые скорости этих двух элементов относительно наблюдателя различаются по величине. Элемент  $x$ , находясь ближе к наблюдателю, будет иметь большую угловую скорость. Такое различие в скорости пытались использовать [Браунштейн, 1962; Гершбергер и Старжец, 1974] для исключения неоднозначности в определении структуры и способа вращения. При этом исходили из того, что более близкие точки отличаются от более далеких большими скоростями. Оценка относительной глубины по различиям угловой скорости, обычно называемая *двигательным параллаксом*, основывается на связи между угловой скоростью и расстоянием. Пусть некоторый элемент перемещается в пространстве со скоростью  $v$  под углом  $\alpha$  к линии прямой видимости и на расстоянии  $r$  от наблюдателя, тогда угловая скорость  $\omega$  этого элемента относительно наблюдателя равна  $\omega = v \sin \alpha / r$ . Говорят, что в некоторой процедуре используется двигательный параллакс, если для определения  $r$  используются  $\omega$  и  $v \sin \alpha$ . Простым примером этого служит однородное смещение окружающей обстановки, вызванное сдвигом наблюдателя. Вдоль заданной линии прямой видимости значение  $v \sin \alpha$  постоянно, и поэтому угловые скорости элементов вдоль этой линии обратно пропорциональны их расстояниям до наблюдателя. Двигательный параллакс использовался для объяснения способности человека ощущать вращение. Далее станет ясно, что применимость двигательного параллакса для объяснения этой способности слишком ограничена, тогда как полярно-параллельная схема представляется подходящей в этой ситуации.

Описанный выше пример вращающегося стержня является более сложным по сравнению с однородным сдвигом. Хотя значения скоростей обоих концов стержня равны (но противоположны по направлению), прежнее соотношение между скоростью и расстоянием не сохраняется. Так как углы между векторами скоростей элементов и соответствующими линиями прямой видимости различны, то разность угловых скоростей оказывается довольно сложной функцией положения стержня [Гершбергер, 1967]. Можно показать, что утверждение «более быстрый конец стержня является более близким» ошибочно для *половины периода вращения*. Если в начальный момент стержень находился на фронтальной плоскости, тогда в пределах поворота на  $45^\circ$  более близкая точка оказывается более медленной. От этого положения до момента, когда стержень виден «головой вперед», более близкая точка становится более быстрой. Угол перехода от медленного к быстрому движению зависит от отношения длины стержня к расстоянию от его средней точки до наблюдателя. По мере увеличения отношения этот угол достигает  $45^\circ$ . Если, например, расстояние до стержня в 2 раза больше его длины, то угол перехода составляет  $50^\circ$  относи-

тельно фронтальной плоскости. В этом случае двигательный параллакс приводил бы к ошибочным результатам почти для 55% периода вращения.

Предшествующий анализ показал, что двигательный параллакс нельзя надежно использовать без оценки ориентации стержня<sup>4,8</sup>. Для оценки преимущества полярно-параллельной схемы перед описанным выше объяснением вместо вращающегося стержня рассмотрим вращающуюся плоскость (чтобы обеспечить полярно-параллельную схему достаточным числом точек). Если возможную ориентацию вращающейся плоскости ограничить углом  $45^\circ$ , где наиболее близкие точки оказываются более медленными, то «параллаксный подход» заведомо приведет к ошибочным результатам, а полярно-параллельная схема даст правильный ответ. Существуют дополнительные серьезные проблемы, связанные с использованием двигательного параллакса. Его нельзя применять, когда стержень вращается не вокруг своей средней точки, когда движущиеся элементы не находятся под углом  $180^\circ$  по отношению друг к другу (т. е. две конечные точки и ось вращения не лежат в одной плоскости) и когда движение стержня не ограничено только чистым вращением.

Другая информация, использовавшаяся для определения направления вращения, мало помогает в тех же случаях, в частности при интерпретации сложного движения. Например, для исключения неоднозначности в определении вращения [Гершбергер и Урбан, 1970] предлагалось использовать неподвижные точки (см.  $p$  и  $q$  на рис. 4.6). Обе эти точки расположены на ближайшей к наблюдателю стороне фронтальной плоскости, поэтому, когда исчезает наблюдаемое движение элемента, его можно «объявить» находящимся между наблюдателем и фронтальной плоскостью. Эта дополнительная информация была бы достаточной для исключения неоднозначности в определении структуры и вращения. Однако этот критерий, как и двигательный параллакс, не работает в случае сложных движений. Поэтому такого рода дополнительную информацию можно использовать в лучшем случае в качестве ненадежных догадок в ситуациях, когда данных для восстановления структуры и вращения с помощью полярно-параллельной схемы оказывается недостаточно.

Непригодность традиционного подхода, основанного на двигательном параллаксе в случае сложного движения, отражает важные различия между ним и схемой интерпретации структуры по движению. В этой схеме разбиение, структура, вращение и сдвиг не рассматриваются независимо (в отличие от работ [Браунштейн, 1962; Гибсон, 1957; Хэй, 1966; Борьессон и Хофстен, 1972, 1973; Йоханссон, 1974; Эрикссон, 1974]). Если имеется достаточный объем информации (достаточное число точек и изображений), то разбиение, структура, вращение и сдвиг определяются однозначно, хотя и без помощи какого-либо отдельного признака.

**Восприятие удаляющихся объектов.** Полярно-параллельную схему можно применить к удаляющимся объектам с целью опре-

деления их структуры и движения. Как будет показано, восстановление структуры оказывается более непосредственным и надежным, чем восстановление сдвига в глубину. На рис. 4.7 изображен объект, движущийся от наблюдателя. Детально этот случай анализироваться не будет. Достаточно отметить, что при движении объекта как целого в глубину различные его части поворачиваются относительно наблюдателя (например, на угол  $\alpha$  на рис. 4.7), и поэтому их структуру можно восстановить. Следо-

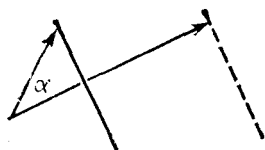
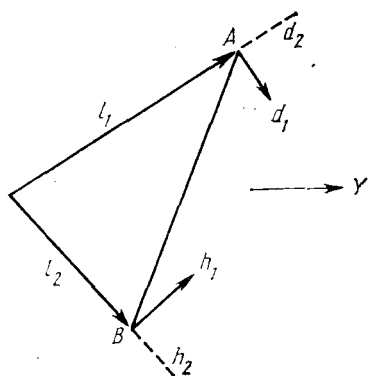


Рис. 4.7. Движение от наблюдателя



Рис. 4.8. Определение сдвига в глубину



вательно, восстановление структуры и вращения (в случае, когда движение от наблюдателя сопровождается вращением), возможно, если объект большой и имеет текстуру, а также если сдвиг в глубину достаточно велик.

Восстановление сдвига в глубину также возможно. Похоже, что среди различных параметров движения сдвиг в глубину в меньшей степени поддается восстановлению с помощью полярно-параллельной схемы. Искажения изображения, вызванные сдвигом в глубину, дают большой объем информации, что делает этот тип движения особенно простым для анализа с помощью прямого перспективного метода. В полярно-параллельной схеме эти искажения рассматриваются как шум, осложняющий вычисления и анализ удаляющихся объектов.

Возможный метод для восстановления сдвига в глубину изображен на рис. 4.8. Пусть элементы  $A$  и  $B$  принадлежат одному и тому же твердому телу  $O$ . Схема интерпретации структуры по движению восстанавливает поворот объекта, обусловленный этим сдвигом, после чего поворот можно уже не рассматривать. Иначе говоря, по двум старым картинкам вычисляются две новые, соответствующие движению этого объекта, которое теперь состоит только из чистого сдвига<sup>4,9</sup>. После исключения поворота из рассмотрения остающееся смещение элементов оказывается только сдвигом (см. рис. 4.8). Остаточный сдвиг элемента  $A$  есть  $d_1$ , причем из-за того, что параллельные проекции являются неоднозначными к сдвигу вдоль линии прямого видения, остающийся

сдвиг вполне может иметь неизвестную компоненту  $d_2$  вдоль линии  $l_1$ . Остаточный сдвиг элемента  $B$  есть  $h_1$ , который, возможно, имеет компоненту  $h_2$  вдоль линии прямого видения  $l_2$ . Двух известных компонент вполне достаточно для определения и истинного движения объекта, даже если оно направлено от наблюдателя (т. е. вдоль  $Y$  на рис. 4.8). Хотя восстановление движения в глубину в принципе реально, оно вовсе не является простым и оказывается чувствительным к ошибкам при определении любого параметра движения. Возможно (это соответствует экспериментальным данным), что восприятие удаляющихся объектов относится в некотором смысле к частному случаю восприятия, основанному на измерении уменьшения размера объекта в целом. В связи с этим необходимо отметить, что в случае удаляющегося трехмерного объекта не происходит равномерного уменьшения размера его перспективного изображения, т. е. изображения различных частей этого объекта могут уменьшаться в разной степени. Однако некоторая степень однородного уменьшения размеров объекта все же может быть достаточной для определения с необходимой точностью эффекта сдвига в глубину.

#### **4.6. Отношение схемы интерпретации структуры по движению к психологии**

В этом разделе обсуждается отношение, которое имеет к психологии схема интерпретации структуры по движению, основанная на предположении о сохранении формы. Здесь приводятся различные экспериментальные данные, наводящие на мысль о том, что принципы, лежащие в основе полярно-параллельной схемы, используются также и зрительной системой человека для вывода структуры из движения. Особое внимание уделяется следующим общим принципам: а) интерпретация выполняется на основе отдельных идентифицируемых элементов; б) глобальная интерпретация объединяет результаты локальных интерпретаций, основанных на сохранении формы, для ядер элементов; в) при интерпретации каждого локального ядра используется приближенные ортогональной проекции.

*Необходимость в идентифицируемых элементах.* Схема интерпретации структуры по движению помогает объяснить явление, которое Уоллок и О'Коннэлл в своей первой работе по *кинетическому эффекту глубины* рассматривали как непостижимую загадку. Когда объекты в этих экспериментах брались с плавно искривленными поверхностями, так что их тени не соответствовали каким-либо идентифицируемым и прослеживаемым трехмерным точкам, то их трехмерную структуру нельзя было определить по проекции: «Криволинейные контуры, деформированные так, что они не содержали явного признака формы, который идентифицировал бы характерную точку на этой кривой, воспринимаются искажающимися [а не движущимися в глубину]... Эта особенность находится в противоречии с нашим объяснением эффекта кинетической



глубины и задержала нашу работу на годы» [Уоллок и О'Коннэлл, 1953, с. 209].

В таких же условиях следует ожидать неудачи с восстановлением структуры по движению и от нашей схемы интерпретации. В соответствии с этой схемой разные изображения не просто связываются в одно целое каким-то способом. Напротив, движение отдельных элементов проверяется на согласованность с движением твердого тела как целого. Следовательно, эта схема интерпретации не будет работать, когда отсутствуют идентифицируемые элементы, которые надежно прослеживаются в ходе двумерных преобразований.

*Число точек.* Характерной чертой схемы интерпретации структуры по движению следует считать то, что структуру можно определить по небольшому числу элементов. Всегда достаточно четырех точек, не лежащих в одной плоскости. Вместе с тем иногда достаточно трех точек, особенно если имеется больше трех изображений. Для зрительной системы человека интерпретация не выглядит процессом, который можно назвать «все или ничего». Точность и стабильность воспринятой структуры увеличивается с числом элементов и изображений [Грин, 1961; Браунштейн, 1962]. Однако необходимый для правильной интерпретации минимум сравним с величиной, требуемой в схеме интерпретации структуры по движению: правильная структура, состоящая всего лишь из трех движущихся элементов, иногда надежно воспринимается.

*Обращение.* Как объяснялось при обсуждении теоремы о структуре по движению, интерпретация в чистом ортогональном случае определяется с точностью до отражения относительно фронтальной плоскости. При восприятии объекта возможно его выворачивание в глубину, которому должно сопутствовать переключение осязаемого вращения. Во многих экспериментах [например, Уоллок и О'Коннэлл, 1953; Уоллок, О'Коннэлл и Нейссер, 1953; Уат и Мюзер, 1960] было установлено, что объекты, наблюдаемые в случае ортогональной проекции, действительно подвергаются спонтанным обращениям в глубину, которым сопутствуют переключения направления осязаемого движения.

*Две движущиеся точки.* Некоторые случаи, когда воспринимаемая структура оказывается неоднозначной, все же удается интерпретировать, если делаются дополнительные предположения. Примером может служить движение одиночного отрезка линии (или двух его концов). Такие две точки всегда можно интерпретировать как концы твердого стержня, ориентация и поворот которого определяются неоднозначно. Однако две недостающие переменные связаны друг с другом: как только становится известной ориентация, определяется поворот, и наоборот. Йоханссон и Янссон [1968] изучали восприятие движущейся одиночной линии. Полученные ими результаты, касающиеся оценки ориентации стержня, свидетельствуют о тенденции наблюдателя считать, что стержень лежит во фронтальной плоскости в тот момент, когда его изображение имеет наибольшую длину.

Заметим, что такое предположение о «наибольшем удлинении», вообще говоря, нельзя включить в схему интерпретации. Вновь рассмотрим пример с двумя цилиндрами и предположим, что наблюдается поворот на угол меньше  $90^\circ$ . В этом случае ориентации пар точек не соответствуют предположению о наибольшем удлинении и воспринимается правильная структура, а не структура, основанная на указанном предположении. (Такой эксперимент был выполнен Уоллоком и О'Коннэллом [1953]. Для обнаружения структуры не показанного полностью объекта достаточно было поворота в пределах  $42^\circ$ .) Предположение о наибольшем удлинении может служить только грубым и ненадежным «последним средством» в случаях, когда общая схема интерпретации (требующая однозначности) не работает. Это предположение, в общем случае ненадежное, оказывается все же наименее рискованным для простой ситуации с двумя элементами. Похоже на то, что зрительной системе человека свойственна тенденция использовать его в таких случаях, но без особенного доверия к результатам: наше трехмерное восприятие вращающегося стержня обычно либо выражено и весьма нестабильно.

*Плоские объекты.* Они, естественно, не удовлетворяют нашим требованиям «непланарности», поэтому не гарантируется правильное определение их структур. Хотя для большинства плоских объектов структура все же будет определяться, некоторым случаям свойственна неоднозначность. Пусть, например,  $l$  будет пересечением плоскости объекта с плоскостью изображения. Предположим также, что объект поворачивается относительно некоторой оси, параллельной  $l$ . В этой ситуации устанавливается факт нахождения точек в одной плоскости, но начальная ориентация плоскости и ее последующие повороты остаются зависимыми, но неизвестными переменными. Дж. Гибсон и Э. Гибсон [1957] обнаружили, что плоские объекты в случае ортогональной проекции действительно воспринимаются неоднозначно при описанном повороте. (По аналогии со случаем с двумя точками в этой недоопределенной ситуации люди, похоже, обнаруживают тенденцию интерпретировать либо начальное, либо соответствующее наибольшему удлинению положение рассматриваемой плоскости как фронтальное.) В отличие от указанного выше случая, структура наклоненной плоскости (не проходящей через вертикальную ось), которая поворачивается относительно вертикальной оси, определяется как схемой интерпретации структуры по движению, так и человеком.

*Абсолютная и относительная глубина.* Схема интерпретации структуры по движению определяет структуру сохраняющих форму объектов. Имеется возможность определить структуру, характеризующуюся относительной глубиной, но нельзя определить абсолютную глубину. Нельзя определить также и относительную глубину двух объектов, движущихся независимо друг от друга. В экспериментах, выполненных Гибсоном и другими [1959], испытанные могли определять истинный наклон движущейся плоскости,

но оценки абсолютного расстояния варьировались от трех дюймов до пяти миль. При предъявлении нескольких плоских объектов в процессе восприятия они разделялись по глубине, когда двигались строго вместе [Гибсон, 1957], но не тогда, когда их движения были независимыми [Гибсон и другие, 1959].

*Влияние размера, текстуры и наклона на единственность решения.* Уже отмечалось, что наиболее предпочтительными условиями для различения правильной интерпретации и ее зеркального образа с помощью полярно-параллельной схемы являются большой угловой размер и мелкая текстура. Обсуждалось также, что для вращающихся плоских объектов выделение единственной интерпретации облегчается, когда плоскость наклонена. Эти ожидания подтверждаются полученными данными в экспериментах с трапециевидным окном Эймса [Эймс, 1951]. Вероятность восприятия правильных ориентаций и вращения этого окна зависит от его размера, текстуры и наклона [Зегерс, 1964; Эпштейн и другие 1968; гл. 5 настоящей книги].

Полярно-параллельная схема уже сравнивалась с двигательным параллаксом при обсуждении возможности получения единственной интерпретации направления вращения. Напомним, что для надежного различия истинной и ложной осей с помощью полярно-параллельной схемы требуется наличие в объекте достаточного числа точек, находящихся на достаточном угловом расстоянии друг от друга. Имеются основания [Уайт и Мюзер, 1960] утверждать, что механизм человеческого восприятия подобен в этом отношении полярно-параллельной схеме. Для небольшого числа элементов воспринимаемое вращение испытывает спонтанные обращения в случае как полярной, так и параллельной проекций.

*Небольшое число далеко разнесенных точек.* В случае полярно-параллельной схемы точность определения трехмерной структуры по перспективным проекциям небольшого числа (около четырех-пяти) точек уменьшается по мере возрастания эффекта перспективных изменений. Последние зависят от отношения между размером объекта и абсолютным расстоянием от него до точки наблюдения. Чем меньше перспективные искажения, тем ближе эта проекция к ортогональной и поэтому выше точность полярно-параллельной схемы. Напротив, для схемы, использующей непосредственно перспективные проекции, значительные перспективные искажения не должны служить помехой для интерпретации. Данные, полученные Бранштейном [1962], наводят на мысль, что сохраняющее форму восприятие объектов зависит от перспективы именно таким образом. Восприятие является наиболее устойчивым для ортогональной проекции, в то время как эффекты перспективы приводят к восприятию искаженного движущегося объекта.

*Определение структуры удаляющихся объектов.* Как было показано, движение от наблюдателя в меньшей степени поддается интерпретации с помощью полярно-параллельной схемы, чем, например, вращение. Наиболее предпочтительные условия для про-

есса интерпретации включают в себя большое угловое различие и мелкую текстуру. Похоже, что такие же требования имеют место в зрительном восприятии человеком структуры удаляющихся объектов. Гибсон и другие [1959] показали, что наклон удаляющейся плоскости хорошо воспринимается в условиях, благоприятных и для полярно-параллельной схемы. Применявшиеся в экспериментах плоские объекты имели плотную текстуру и угловой размер, равный  $82^\circ$ . Когда полярно-параллельный алгоритм применялся к той же удаляющейся плоскости, то ее структура определялась правильно и единственным образом. Уже отмечалось, что в примере двух удаляющихся цилиндров (см. подразд. 4.5.1), где условия были менее благоприятными с точки зрения полярно-параллельной схемы (меньший угловой размер и меньшее число элементов), восприятие структуры по движению от наблюдателя казалось невозможным, в отличие от вращения, которое допускало интерпретацию.

*Движение, не сохраняющее форму.* Схему определения структуры по движению нельзя применить к деформациям, при которых не сохраняется форма объектов. Однако поскольку процесс интерпретации является локальным и требует лишь небольшого числа изображений, то эта схема применима к движениям, почти сохраняющим форму, которые локально (в смысле пространства и времени) аппроксимируются движением, сохраняющим форму (например, движение, соответствующее изгибу [Янссон и Йоханссон, 1973]).

## ГЛАВА 5

### ВОСПРИЯТИЕ ДВИЖЕНИЯ ПО СТРУКТУРЕ

Предыдущая глава была посвящена проблеме построения описаний трехмерной структуры и движения исходя из двухмерных преобразований, когда каждое статическое изображение сцены не содержит явной трехмерной информации. Однако даже единственное монокулярное изображение сцены часто передает подробную информацию о трехмерной структуре этой сцены. Здесь не будет обсуждаться то, как возникает такая статическая интерпретация трехмерной структуры. Когда же такая интерпретация действительно имеется, ее можно использовать для определения пространственного движения. В настоящей главе исследуются основные результаты использования статического трехмерного восприятия для интерпретации видимого движения. Если процесс интерпретации, рассмотренный в предыдущей главе, можно сказать, вычисляет структуру по движению (С.п.Д.), то предметом обсуждения в настоящей главе является определение движения по структуре (Д.п.С.), поскольку для восстановления пространственного движения используется знание структуры. Строго говоря,

первый метод строит описания трехмерной структуры и движения в пространстве на основе двумерных трансформаций изображения, в то время как второй выводит движение в пространстве из трансформаций такой интерпретированной трехмерной структуры.

В разд. 5.1 будут приведены примеры, иллюстрирующие существование процесса определения движения по структуре и некоторые его свойства. Схема Д.п.С. требует, чтобы трехмерная структура сцены воспринималась по статическим изображениям, в то время как схема С.п.Д. может работать независимо от наличия статического трехмерного восприятия. Следовательно, когда статическое трехмерное восприятие присутствует, обе схемы могли бы работать одновременно. При их совместной работе возникает проблема объединения результатов интерпретаций, полученных от разных источников. В разд. 5.2 в качестве основы для обсуждения различных вопросов, касающихся объединения С.п.Д. и Д.п.С. схем интерпретаций, используется так называемая иллюзия Эймса.

### 5.1. Схема интерпретации движения по структуре

Статическое восприятие трехмерной структуры может оказать очевидную помощь в интерпретации пространственного движения. Движение объекта представляет собой непрерывное изменение его местонахождения и расположения. Поэтому информация о трехмерной структуре объектов и изменениях структуры во времени непосредственно имеет отношение к движению рассматриваемых объектов. Однако *a priori* не ясно, до какой степени статическая и динамическая интерпретация взаимодействуют друг с другом. Можно представить себе, например, что используются различные внутренние представления, так что статическая трехмерная интерпретация не может быть непосредственно «вписана» в динамическое представление и не может оказать воздействие на восприятие движения, или же что статическая интерпретация осуществляется слишком медленно, чтобы повлиять на восприятие движения. В настоящем разделе будет предпринята попытка показать, что статическая трехмерная интерпретация все же используется в восприятии движения и что статическая и динамическая интерпретации, вероятно, объединяются в некоторое единое представление. Для этого недостаточно будет показать, что можно воспринимать пространственное движение статически интерпретируемых структур. Поскольку С.п.Д. схема помогает восприятию движения в пространстве независимо от того, воспринимаема ли статически трехмерная структура, то потребуются демонстрации, в которых применение схемы интерпретации типа С.п.Д. невозможно. Если С.п.Д. схема не применима, но пространственное движение все же можно оценить и оно согласуется с изменениями воспринятой структуры, то существование Д.п.С. схемы было бы в значительной степени подтверждено.

В этом разделе рассматриваются специально подготовленные демонстрации, которые не допускают С.п.Д. интерпретацию, что делается путем нарушения различных условий, существенных для успешного применения С.п.Д. схемы. Эту схему можно исключить различными способами: задавая недостаточное число точек (пример 1), недостаточное число изображений (пример 2), не сохраняющее форму движения (примеры 3 и 4) или проекцию некоторых плоских объектов (таких, как окно Эймса, описанное в разд. 5.2). Во всех этих случаях может возникнуть восприятие непрерывного движения в трехмерном пространстве при условии, что указанная трехмерная структура воспринимается по отдельным изображениям.

**Пример 1. Восприятие движения одиночного отрезка.** Первая демонстрация изображена на рис. 5.1. Отрезки  $l$  (на левой стороне куба) и  $r$  того же куба показываются в ситуации кажущегося движения. Оба отрезка видны в течение всего времени предъявления. Однако на первой картинке отрезок  $l$  — яркий, а отрезок  $r$  — тусклый, в то время как на второй картинке отрезок  $r$  — яркий, а отрезок  $l$  — тусклый. Когда две картинке показывают поочередно, отрезки  $l$  и  $r$  воспринимаются участвующими в кажущемся движении, а трехмерная структура этого куба воспринимается в течение всего времени предъявления. При кажущемся движении отрезка от  $l$  к  $r$  оно воспринимается как непрерывное перемещение в глубину. Зрительная система «заполняет» траекторию этого отрезка, и такое заполнение имеет место в трехмерном перцептивном пространстве. Начальное и конечное положения траектории определяются в этом случае с помощью статической трехмерной интерпретации, а не С.п.Д. схемой. В отличие от движения изолированной линии [Йохансон и Янссон, 1968], в этом примере позиция, обозначенная символом  $l$ , в которой длина отрезка принимает свое максимальное значение, не интерпретируется как отрезок, лежащий во фронтальной плоскости. Восприятие пространственного движения, пожалуй, согласуется с имеющейся статической трехмерной интерпретацией этой сцены. Более того, когда куб выворачивается в глубину, то соответствующим образом изменяется и воспринятое движение этой линии. Восприятие пространственного движения в таком случае можно описать, говоря, что зрительная система приписывает отрезкам  $l$  и  $r$  различные трехмерные метки (термин «трехмерная метка» будет использоваться в перцептивном пространстве как краткое обозначение длины и трехмерной ориентации, приписанных линии, или ориентации, приписанной поверхности). Когда процесс нахождения соответствия идентифицирует указанные два отрезка, как, в сущности, одно и то же, изменение в их трехмерных метках интерпретируется как движение в пространстве, траектория движения заполняется и отрезок затем воспринимается движущимся непрерывно от начального к конечному положению.

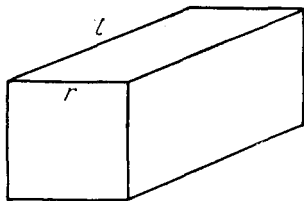


Рис. 5.1. Отрезки  $l$  и  $r$  воспринимаются участвующими в кажущемся движении

Итак, статическое восприятие трехмерной структуры может привести к восприятию непрерывного движения в трехмерном пространстве. Это имеет место, когда траектория между элементами

с различными (статическими) трехмерными метками достраивается зрительной системой. Эти два элемента должны находиться в определенном отношении соответствия для того, чтобы упомянутое заполнение имело место и чтобы такое движение воспринималось. Например, изменения трехмерных меток при обращениях куба Неккера не сопровождаются восприятием движения.

**Некоторые выводы, касающиеся представления зрительной информации.** Как утверждалось во введении, вопросом первостепенной важности для вычислительного подхода к исследованиям зрительного восприятия является природа внутреннего представления зрительной информации. Теперь можно суммировать некоторые общие соображения, которые возникают при изучении процессов нахождения соответствия и определения структуры по движению и движения по структуре относительно вида представления, связанного с видимым движением. Имеются данные о существовании, по крайней мере, двух различных уровней представлений, где каждый уровень объединяет зрительную информацию, получаемую от ряда источников (ср. [Марр, 1977b]). Первым уровнем является пространство опорных символов, которое двумерно (т. е. на этом уровне не представлена глубина) и в котором реализуются вычисления соответствия. Как было показано в гл. 2, опорные символы в этом пространстве берутся из различных источников, таких как изменения интенсивности в изображении, стереоразличия, текстурные градиенты и опорные символы групп. Второй уровень является трехмерным пространством, где статическая и динамическая интерпретации объединяются, а траектории заполняются при восприятии.

Существование таких различных представлений могло бы служить объяснением очевидного несоответствия между результатами, изложенными в разд. 2.5, и более ранними экспериментами [Корбэн, 1942; Эттнив и Блок, 1973]. В экспериментах, описанных в разд. 2.5, опробовалась метрика сродства, которая связана с пространством опорных символов и поэтому зависит от их двумерного разделения. В отличие от разд. 2.5, в котором рассматривались методы, основанные на конкурирующих движениях, эти более ранние эксперименты основывались на оценке гладкости движения. Суждения о гладкости, вероятно, в основном зависят от процесса заполнения траекторий, который имеет место в трехмерном пространстве представления.

**Пример 2. Восприятие движения по двум параллельным проекциям.** Пространственное движение может быть воспринято всего лишь по двум параллельным проекциям, если начальная и конечная трехмерные структуры поддаются оценке. Когда, например, рисунки, изображающие параллельную проекцию вращающегося куба, показываются так, что возникает кажущееся движение, тогда пространственное вращение куба воспринимается по двум картинкам. Такое восприятие следует противопоставить С. п. Д. схеме интерпретации, для которой структура и движение остаются неоднозначными, если только не предьявляются три различные изображения. Таким образом, С. п. Д. схема не может

объяснить такую интерпретацию. Наблюдаемое движение согласуется с изменениями статических трехмерных меток, приписанных линиям и поверхностям фигур. Следовательно, тот же процесс, описанный в примере 1, будет пригодным для проведения интерпретации и в этом случае.

**Пример 3. Восприятие движения, в котором форма объектов не сохраняется.** Поскольку С.п.Д. схема настроена на интерпретации, сохраняющие форму, интересно рассмотреть восприятие движений, не сохраняющих форму. Если по статическим изображениям воспринимаются начальная и конечная трехмерные структуры, то движение в пространстве воспринимается даже в том случае, если оно вызывает превращения, не сохраняющие форму. Например, фигуры различной формы, ограниченные плоскостями, такие как пирамида и куб, можно показывать попеременно, что приводит к ситуации кажущегося движения. Можно увидеть, как пирамида гладко превращается в куб и обратно в пирамиду. При такой деформации различные ребра движутся независимо друг от друга, в результате чего возникает не сохраняющее форму движение, которое не может быть определено С.п.Д. схемой.

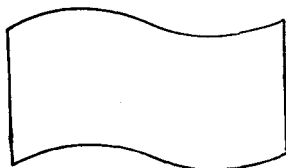


Рис. 5.2. Развевающийся флаг

Другой пример состоит в демонстрации флага (рис. 5.2). Этот флаг можно было бы видеть развевающимся и испытывающим не сохраняющие форму искажения. Искажения флага удавалось оценить лишь по двум изображениям.

Если даны С.п.Д. и Д.п.С. схемы интерпретации, то движение, не сохраняющее форму, можно определить одним из двух способов. Как отмечалось в гл. 4, если некоторое движение локально аппроксимируется (как в пространстве, так и во времени) превращением, сохраняющим форму, то его можно обрабатывать С.п.Д. схемой. Такую интерпретацию можно было бы применить, например, для анализа изображения идущего человека. Фигура человека в целом (обычно) подвергается превращению, не сохраняющему форму. Однако, когда рассматриваются небольшие части его тела, их движение приблизительно сохраняет форму, поэтому С.п.Д. схему можно применить для определения их структуры и движения. Другие случаи, как, например, превращения пирамида—куб или развевающийся флаг, не удастся интерпретировать с помощью С.п.Д. схемы, но можно обработать Д.п.С. схемой при условии, что начальная и конечная трехмерные структуры могут быть восприняты. Если бы обе указанные схемы были единственно доступными для интерпретации не сохраняющих форму превращений, то можно было бы предсказать, что не сохраняющие форму движения, статические проекции которых не приводят к трехмерному восприятию, не привели бы и к ясной и точной интерпретации структуры и движения в пространстве. Все же в отдельных специальных ситуациях может возникать некоторое впечатление пространственного движения. В частности, уменьшение размера изображения, как правило, приводит к ощущению движения от наблюдателя, даже когда определить трехмерную структуру не удастся (см. подразд. 4.5.2).

**Пример 4. Иллюзия Маха.** Иллюзию Маха можно продемонстрировать следующим образом. Рассмотрим согнутый лист бумаги, образующий стоящую v-образную фигуру. При монокулярном наблюдении эта фигура воспринимается неоднозначно — она может выворачиваться в глубину [Иден, 1962; Линдсэй и



Норман, 1972]. Нужно смотреть на эту фигуру, закрыв один глаз, а другой оставив открытым (монокулярно), и ждать появления эффекта выворачивания. Если такой эффект возникает, то можно медленно двигать головой вправо—влево, вверх—вниз или вперед—назад. Результат поразительный: кажется, что объект перемещается вместе с движением головы. (Похожие иллюзии можно получить и с другими конструкциями, например с проволочным кубом.) Такое иллюзорное движение появляется, несмотря на знание наблюдателем истинной ситуации, и часто противоречит критериям стабильности, информации о тени и о тактильных ощущениях [Иден, 1962]. Уместно заметить, что при положении объекта на небольшом расстоянии от глаза его движение сопровождается значительными искажениями — оно более не сохраняет форму. Следовательно, иллюзорное движение в этом случае нельзя отнести к возникающей при выворачивании неоднозначности С.п.Д. схемы. Если полярно-параллельная схема была бы применена к такой конфигурации, то она обнаружила бы только правильную структуру. Можно сказать, что эта схема невосприимчива к иллюзии Маха. Скорее всего люди, которым предъявляется конфигурация Маха, состоящая только из несвязанных элементов и без видимых контуров, будут также менее восприимчивы к иллюзии Маха. Все же простые выворачивания могли бы иметь место, но в этом случае искажения, типичные для иллюзии Маха, исчезнут.

Дополнительное подтверждение того, что неправильное восприятие структуры вызывает ошибочную интерпретацию движения, будет приведено в начале разд. 5.2 при обсуждении иллюзии Эймса.

Применение к некоторой сцене С.п.Д. и Д.п.С. схем обычно не приводит к противоречивым результатам. Этого, однако, не произойдет, если оказывается ошибочным статическое восприятие структуры, которое управляет Д.п.С. процессом. Когда возникает такая двусмысленная ситуация, зрительная система сталкивается с проблемой согласования конфликтующих источников информации. Существуют две основные возможности для разрешения таких конфликтов: соглашение о некотором среднем решении или выбор одного решения и исключение другого. Действительно, при сильном конфликте С.п.Д. схемы и статического трехмерного восприятия выбирается только одно из решений<sup>5.1</sup>. Следовательно, при восприятии в зависимости от конкретной ситуации иногда полностью преобладает один из процессов (с возможными переключениями между ними). Иллюзия Маха является примером доминирования Д.п.С. схемы. В свою очередь Уоллок, О'Коннэлл и Нейссер [1953] привели примеры, в которых тень от проволочных объектов восстанавливалась как трехмерная структура, не конгруэнтная реальной. Реальная структура обычно обнаруживалась лишь тогда, когда начинали передвигать эти объекты. В этом примере С.п.Д. процесс преобладал над Д.п.С. схемой.

Разнообразно изучавшаяся иллюзия окна Эймса дает пример соперничества этих процессов [Эймс, 1951]. В следующем разделе рассматривается иллюзия Эймса и показывается, что неко-

которые ее основные черты проявляются, если для объяснения использовать соперничество между двумя схемами интерпретации. Обсуждение иллюзии Эймса представляет интерес также и потому, что позволит объединить различные результаты, обсуждающиеся в гл. 4 и 5.

## 5.2. Феномены, связанные с иллюзией Эймса, и соперничество между С.п.Д. и Д.п.С. схемами

Окно Эймса представляет собой плоскую трапециевидную конструкцию, выглядящую как перспективное изображение прямоугольного окна. Когда окно вращается относительно вертикальной оси, то вместо вращения часто воспринимается его колебательное движение [Эймс, 1951]. Эта иллюзия не ограничивается только конструкцией в виде окна. Некоторые другие фигуры при вращении также выглядят колеблющимися, в то время как на самом деле они участвуют в непрерывном вращении. Возможность неправильного восприятия направления вращения окна сама по себе не удивительна. Легко можно увидеть, что если отрезок  $g$  на рис. 5.3 воспринимается расположением дальше, чем отрезок  $c$  (оба в действительности лежат в фронтальной плоскости), то при вращении трапеции относительно  $c$  в *любом* направлении единственной согласующейся интерпретацией будет вращение против часовой стрелки (если смотреть сверху). Как и в случае иллюзии Маха, неправильные суждения о структуре приводят к неправильному восприятию движения.

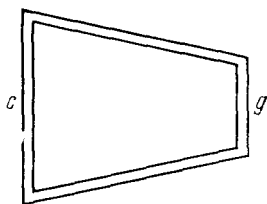


Рис. 5.3. Трапециевидное окно

Однако, в отличие от иллюзии Маха, неправильно воспринятая структура окна Эймса не совпадает с отражением истинной структуры относительно плоскости изображения. Если бы иллюзия Маха имела место в условиях параллельной проекции, то она объяснялась бы без привлечения Д.п.С. процесса, как, например, просто в случае восприятия отраженной, а не реальной структуры. Тем не менее иллюзию Эймса можно использовать для демонстрации влияния Д.п.С. процесса даже в условиях параллельной проекции, поскольку иллюзорное движение нельзя объяснить двужначностью, свойственной этой проекции.

В большинстве работ утверждалось, что собственно иллюзия Эймса является результатом неправильного восприятия структуры. Однако менее очевидными можно считать причины, из-за которых эта иллюзия небезупречна, иначе говоря, не совсем ясно, что выдает истинное положение фигуры. При различных условиях эта иллюзия исчезает и воспринимается истинное вращение в течение значительной части времени наблюдения. Бинокулярное зрение является одним из очевидных источников информации, говорящей о реальной действительности, однако реальная структура

и вращение часто воспринимаются и при монокулярном наблюдении. В качестве возможных объяснений предлагались подсказки типа двигательного параллакса, но вследствие изложенных в подразд. 4.5.2 причин они могут объяснить правильное восприятие только в ограниченных условиях. Очевидным кандидатом для получения правильной интерпретации является С.п.Д. схема, которой несвойственны неправильные восприятия структуры. Феномен Эймса, таким образом, можно рассматривать как соревнование двух процессов, а именно Д.п.С. схемы, ответственной за ошибочную интерпретацию, и С.п.Д. схемы, обеспечивающей правильную интерпретацию. В условиях, неблагоприятных для С.п.Д. схемы. *Наклоненная плоскость* представляет собой нетерптерпаций, в то время как при благоприятных условиях можно ожидать ослабления иллюзии<sup>5.2</sup>.

Ниже будут рассмотрены пять феноменов, связанных с иллюзией Эймса, для трех вариантов наблюдения. Относительно этих феноменов будут сделаны определенные предсказания и по возможности сопоставлены с эмпирическими данными. Тремя вариантами наблюдения являются следующие: 1) вертикальная плоскость в условиях параллельной проекции; 2) наклоненная плоскость в условиях параллельной проекции; 3) вертикальная плоскость в условиях перспективной проекции. *Вертикальная плоскость* означает, что окно проходит через вертикальную ось и вращается относительно этой оси. Это — пример конфигурации, интерпретация которой оказывается неоднозначной для параллельной С.п.Д. схемы. *Наклоненная плоскость* представляет собой нефронтальную плоскость, которая проходит через горизонтальную ось, но вращается относительно вертикальной оси. Структура такой плоскости определяется параллельной С.п.Д. схемой (см. подразд. 4.2.5).

Эти три варианта наблюдения имеют существенные для последующего обсуждения различия. Основное различие заключается в применимости параллельной и полярно-параллельной С.п.Д. схем. В случае первого варианта обе схемы не работают. Вторым и третьим варианты поддаются анализу только с помощью параллельной и полярно-параллельной схем соответственно. Ниже исследуются более подробно применимость С.п.Д. и Д.п.С. схем к указанным вариантам. Затем эти различия будут использованы для анализа феноменов, связанных с иллюзией Эймса.

Параллельная С.п.Д. схема интерпретации применима ко второму варианту, а полярно-параллельная — к третьему. К первому варианту ни параллельная, ни полярно-параллельная схема неприменимы. Следовательно, можно ожидать, что в этом случае Д.п.С. схема будет преобладать. Отказ параллельной С.п.Д. схемы является следствием планарности объекта. Неплоские объекты относятся ко второму варианту, т. е. можно ожидать, что они будут определены С.п.Д. схемой неоднозначно с точностью до отражения. Второй вариант поддается интерпретации параллельной С.п.Д. схемой, но поскольку полярно-параллельная схем

ма здесь неприменима, структуру можно определить только с точностью до отражения. Для третьего варианта структуру можно определить *единственным* образом с помощью полярно-параллельной С.п.Д. схемы. С увеличением углового размера, наклона и текстуры (плотности характерных точек) облегчается интерпретация с помощью С.п.Д. схемы, и поэтому можно предсказать их воздействие на рассматриваемые ниже феномены. Предсказания будут иметь качественный характер, поскольку не учитываются некоторые вторичные эффекты<sup>5,3</sup> и относительные эффективности обсуждаемых двух схем интерпретации можно сравнить только качественно.

**Анализ пяти феноменов, связанных с иллюзией Эймса.** Рассматриваемые ниже феномены включают в себя обращение фигуры, предпочтительные положения обращения, предпочтительную зону ориентации, предпочтительное ощущение направления вращения и определение правильной ориентации.

*Феномен 1: обращения направления вращения.* Некоторые недоразумения, относящиеся к иллюзии Эймса, возникают при изучении восприятия обращений независимо от других феноменов. Объяснение самим Эймсом этой иллюзии как неправильно воспринятой структуры было подвергнуто сомнению [Пауэр и Дэй, 1973] на основе того, что фигуры, структура которых правильно воспринимается, также подвержены спонтанным обращением направления воспринимаемого вращения (особенно, когда условия наблюдения приближаются к параллельной проекции). При обсуждении С.п.Д. интерпретации отмечалось, что структура и движение *любого* объекта определяются его параллельными проекциями только с точностью до отражения и поэтому оказываются склонными к спонтанным обращениям как структуры, так и направления вращения. В отличие от спонтанных обращений, характеризующих параллельную проекцию, можно ожидать, что обращения в иллюзии Эймса, вызванные неправильным восприятием структуры, имеют место для определенных ориентаций—когда вращающееся окно пересекает фронтальную плоскость. Другое различие между этими двумя обращениями заключается в том, что только со спонтанными обращениями связаны обращения структуры. *Просто появление обращений не следует смешивать с иллюзией Эймса*, и такой путаницы можно избежать, если учитывать также и другие обсуждаемые феномены, например предпочтение положениям обращений. Можно ожидать, что при правильном восприятии структуры для указанных выше первого и второго вариантов будут наблюдаться спонтанные обращения даже в случае неплоских объектов [Уоллок и О'Коннэл, 1953]. Третий вариант должен быть относительно устойчивым к таким обращениям, особенно когда наблюдаемый объект относительно большой и имеет плотно распределенные характерные точки [Гибсон и другие, 1959]. Мало что можно сказать о частоте обращений, хотя она является одним из параметров, оцениваемых в большинстве экспериментов. Причина заключается в том, что час-

тоту спонтанных обращений в условиях параллельной проекции нельзя предсказать, так как она зависит от неизвестных параметров зрительной системы, а не только от рассматриваемых стимулов. Однако можно ожидать, что, когда полностью преобладает Д.п.С. схема, как в первом варианте, частота обращений за период будет приближаться к 2. В экспериментах [Эпштейн и другие, 1968; Кросс и Кросс, 1969] было установлено, что частота спонтанных обращений для С.п.Д. схемы меньше, а именно в экспериментальных условиях она была около 0,5 за период. Поскольку частота обращений в Д.п.С. схеме значительно выше, можно ожидать, что она будет увеличиваться по мере уменьшения роли интерпретации по схеме С.п.Д. по сравнению с интерпретацией по схеме Д.п.С. Эта частота должна быть самой высокой в первом и самой низкой в третьем варианте и уменьшаться в зависимости от размера, наклона и текстуры объекта. Эти предположения находятся в соответствии с экспериментальными данными [Зегерс, 1964; Эпштейн и другие, 1968].

*Феномен 2. Предпочтительные положения обращения.* Можно ожидать, что в первом варианте обращения имеют место всякий раз, когда объект пересекает фронтальную плоскость. Во втором варианте число таких обращений может уменьшаться с возможным появлением спонтанных обращений при других ориентациях. В третьем варианте, возможно, дополнительно будут наблюдаться в произвольных положениях некоторые обращения, обусловленные переключениями от одного вида интерпретации к другому. В отличие от второго варианта, этим спонтанным переключениям сопутствовало бы изменение воспринятой структуры, которая являлась бы необязательно отражением относительно плоскости изображения. Итак, существуют три различных типа обращения при восприятии направления вращения:

1. Спонтанные обращения, связанные с параллельными проекциями, имеющие место в обычных положениях, с сопутствующими обращениями структуры; можно ожидать, что они наиболее часто встречаются во втором варианте.

2. Обращения, обусловленные неправильно воспринятой структурой, имеющие место в определенных положениях и без сопутствующих обращений структуры; можно ожидать, что они имеют место чаще в первом варианте.

3. Обращения, вызванные переходом от одной доминирующей интерпретации к другой, с сопутствующими переходам изменениями структуры, но не обязательно обращениями; можно ожидать, что они имеют место, в основном, в третьем варианте, хотя возможны во втором и исключены в первом вариантах. Эти различия остаются только предположениями для будущих исследований, поскольку изменения структуры, связанные с обращениями, не изучались.

*Феномен 3. Предпочтительная зона ориентации.* Можно ожидать, что в первом варианте ориентация объекта ограничивается

некоторым сектором, определяемым главным образом воспринятой структурой. Если, например, трапециевидное окно, расположенное на фронтальной плоскости, воспринимается как повернутое на  $45^\circ$  относительно фронтальной плоскости, то можно ожидать, что окно будет восприниматься колеблющимся в пределах ориентаций примерно от  $+45^\circ$  до  $-45^\circ$ , охватывая тем самым сектор в  $90^\circ$ . Если бы подлинное движение воспринималось всегда, то отсутствовала бы какая-либо предпочтительная зона ориентации и объект был бы виден во всей зоне, равной  $360^\circ$ . Во втором варианте ожидаемая зона ориентации будет шире, чем в первом, поскольку С.п.Д. схема дает информацию об истинной ориентации. Эти предположения подтверждаются в работе [Олсон, 1974]. Понятно, что из двух ориентаций, связанных с параллельной проекцией, предпочтительной будет та, которая окажется конгруэнтной статической структуре. В результате во втором варианте один сектор, равный  $180^\circ$ , будет предпочтительней другого. Третий вариант должен показать меньшее предпочтение какому-либо определенному сектору, особенно если окно большого размера имеет наклон и текстуру.

*Феномен 4. Предпочтительное ощущение направления вращения.* Схема Д.п.С. привела бы к колеблющемуся движению, в котором не отдавалось бы предпочтение ни одному из направлений вращения. Параллельная С.п.Д. схема также не дает характерного предпочтения определенному вращению. Хотя тип движения является разным в первом и втором вариантах (в первом оно колебательного характера), тем не менее в обоих вариантах отсутствует предпочтение направлению вращения, так же как и для параллельной проекции вообще. В свою очередь третий вариант должен показать предпочтение действительному направлению вращения, поскольку полярно-параллельная схема способна однозначно определять истинное вращение.

*Феномен 5. Восприятие правильной ориентации.* В первом варианте вообще нельзя определить правильную ориентацию. Во втором варианте можно ожидать, что воспринятая ориентация в среднем ближе к действительной или к ее отражению (относительно плоскости изображения). Только в третьем варианте ориентацию можно определить правильно и однозначно. Это подтверждается экспериментальными данными [Гибсон и другие, 1959; Уоллок и О'Коннэлл, 1959].

Итак, феномены, связанные с иллюзией Эймса, можно объяснить с помощью соперничества между С.п.Д. и Д.п.С. схемами, и эти феномены характеризуются следующими тремя принципами: 1) неправильно воспринятая структура приводит к иллюзорным суждениям о структуре и вращении, что объясняется Д.п.С. схемой; 2) полярно-параллельная схема является основой для надежного восприятия; 3) в случаях, сходных с параллельной проекцией, полярно-параллельная схема вырождается в параллельную, которая может дать правильную информацию с точностью лишь до отражения относительно фронтальной плоскости.

## Заклучение

Некоторые важные принципы, лежащие в основе интерпретации видимого движения, можно коротко изложить с помощью трех типов отношений, которые зрительная система устанавливает между парами стимулов. Первым типом является отношение *замещения*. Один объект может восприниматься как исчезающий и замещаемый другим (как и в игре реек-а-воо\*, в которой играющий замещается препятствием). Вторым типом является отношение *преобразования*. Одна фигура может восприниматься как бы превращающейся в другую без потери своей идентичности (как, например, превращение в мультфильме лягушки в царевну). Третьим типом является отношение *движения*, при котором разные изображения воспринимаются как представляющие тот же самый, не меняющийся движущийся объект.

Подходящим заключением для этой работы будет объяснение указанных фундаментальных отношений с помощью процессов нахождения соответствия и интерпретации.

Два предъявленных последовательно стимула в зависимости от пространственно-временных отношений и отношений сродства между их составными частями могут или не могут быть сопоставлены с помощью процесса нахождения соответствия. Ситуации, когда их не удастся сопоставить, приводят к отношению замещения. Для двух соответствующих друг другу стимулов процесс интерпретации может установить различие между изменением структуры и ситуацией движения. Если меняющуюся сцену можно интерпретировать как движущееся твердое тело, то эта сцена так и будет воспринята. В противном случае возникает указание на изменение структуры.

### Приложение 1. Теорема о структуре по движению

**Теорема о структуре по движению:** *если даны три различные ортогональные проекции\*\* четырех точек, не лежащих в одной плоскости и принадлежащих сохраняющей форму конфигурации, то их структура и движение, совместимые с указанными тремя изображениями, определяются однозначно с точностью до симметрии относительно плоскости изображения.*

**З а м е ч а н и е.** Предполагается, что соответствие между проекциями уже установлено. Иначе говоря, известно, какие точки заданных трех изображений являются проекциями той же самой исходной пространственной точки.

**Обозначения.** Пусть  $O, A, B$  и  $C$  — указанные четыре точки. Движение объекта состоит из сдвига и поворота. По ортогональной проекции нельзя восстановить сдвиг в глубину, а остающаяся компонента сдвига восстанавливается тривиально, поскольку она одинакова в пространстве и на плоскости изображения. Предполагается также, что уже найдены соответствующие друг другу точки (т. е. три проекции той же самой пространственной точки). Исходную задачу поэтому можно сформулировать следующим образом: заданы ортогональные проекции четырех точек  $O, A, B$  и  $C$  на три различные плоскости  $\pi_1,$

\* Вид детской игры, распространенной в США, аналогичной игре в жмурки. — Прим. пер.

\*\* Имеются в виду ортогональные проекции на плоскость изображения. — Прим. пер.

$\pi_2$  и  $\pi_3$ . Требуется реконструировать исходную трехмерную конфигурацию этих точек.

Выберем фиксированную систему координат с началом в точке  $O$ . Пусть  $\mathbf{a}$ ,  $\mathbf{b}$  и  $\mathbf{c}$  — векторы от точки  $O$  соответственно к точкам  $A$ ,  $B$  и  $C$ . Пусть каждое изображение имеет двумерную систему координат  $(p_i, q_i)$  с началом в точке, являющейся изображением точки  $O$ , т. е.  $\mathbf{p}_i$  и  $\mathbf{q}_i$  — ортогональные единичные векторы на плоскости  $\pi_i$ . Пусть координаты точек  $A$ ,  $B$  и  $C$  на плоскости  $\pi_i$  (координаты их изображений) есть  $(x_{ai}, y_{ai}, x_{bi}, y_{bi}, x_{ci}, y_{ci})$  для  $i=1, 2, 3$ . Наконец, пусть  $\mathbf{u}_{ij}$  — единичный вектор, принадлежащий линии пересечения плоскостей  $\pi_i$  и  $\pi_j$ .

Координаты изображений рассматриваемых точек задаются:

$$\begin{aligned} x_{ai} &= \mathbf{a} \mathbf{p}_i; & y_{ai} &= \mathbf{a} \mathbf{q}_i; \\ x_{bi} &= \mathbf{b} \mathbf{p}_i; & y_{bi} &= \mathbf{b} \mathbf{q}_i; \\ x_{ci} &= \mathbf{c} \mathbf{p}_i; & y_{ci} &= \mathbf{c} \mathbf{q}_i. \end{aligned} \quad (\text{П1.1})$$

Единичный вектор  $\mathbf{u}_{ij}$  лежит на плоскостях  $\pi_i$  и  $\pi_j$ , точки которых определяются соответственно системами координат  $(p_i, q_i)$  и  $(p_j, q_j)$ . Отсюда

$$\mathbf{u}_{ij} = \alpha_{ij} \mathbf{p}_i + \beta_{ij} \mathbf{q}_i, \quad \alpha_{ij}^2 + \beta_{ij}^2 = 1; \quad (\text{П1.2})$$

$$\mathbf{u}_{ij} = \gamma_{ij} \mathbf{p}_j + \delta_{ij} \mathbf{q}_j, \quad \gamma_{ij}^2 + \delta_{ij}^2 = 1. \quad (\text{П1.3})$$

Из (П1.2) и (П1.3) следует векторное уравнение

$$\alpha_{ij} \mathbf{p}_i + \beta_{ij} \mathbf{q}_i = \gamma_{ij} \mathbf{p}_j + \delta_{ij} \mathbf{q}_j. \quad (\text{П1.4})$$

Скалярно умножая уравнение (П1.4) соответственно на векторы  $\mathbf{a}$ ,  $\mathbf{b}$  и  $\mathbf{c}$ , получаем

$$\begin{aligned} \alpha_{ij} x_{ai} + \beta_{ij} y_{ai} &= \gamma_{ij} x_{aj} + \delta_{ij} y_{aj}; \\ \alpha_{ij} x_{bi} + \beta_{ij} y_{bi} &= \gamma_{ij} x_{bj} + \delta_{ij} y_{bj}; \\ \alpha_{ij} x_{ci} + \beta_{ij} y_{ci} &= \gamma_{ij} x_{cj} + \delta_{ij} y_{cj}. \end{aligned} \quad (\text{П1.5})$$

Эти три уравнения относительно  $(\alpha_{ij}, \beta_{ij}, \gamma_{ij}, \delta_{ij})$  линейно независимы. Чтобы это показать, предположим, что существуют три скаляра  $\eta$ ,  $\lambda$  и  $\mu$ , такие, что

$$\begin{aligned} \eta \mathbf{p}_i \mathbf{a} + \lambda \mathbf{p}_i \mathbf{b} + \mu \mathbf{p}_i \mathbf{c} &= 0; \\ \eta \mathbf{q}_i \mathbf{a} + \lambda \mathbf{q}_i \mathbf{b} + \mu \mathbf{q}_i \mathbf{c} &= 0; \\ \mu \mathbf{p}_j \mathbf{a} + \lambda \mathbf{p}_j \mathbf{b} + \mu \mathbf{p}_j \mathbf{c} &= 0; \\ \mu \mathbf{q}_j \mathbf{a} + \lambda \mathbf{q}_j \mathbf{b} + \mu \mathbf{q}_j \mathbf{c} &= 0. \end{aligned} \quad (\text{П1.6})$$

Отсюда следует, что вектор  $\boldsymbol{\theta} = \eta \mathbf{a} + \lambda \mathbf{b} + \mu \mathbf{c}$  должен быть ортогональным к единичным векторам  $\mathbf{p}_i$ ,  $\mathbf{q}_i$  и  $\mathbf{p}_j$ ,  $\mathbf{q}_j$ . Но поскольку плоскости  $\pi_i$  и  $\pi_j$  различны, то  $\boldsymbol{\theta} = 0$ . Заметим, что точки  $O$ ,  $A$ ,  $B$  и  $C$  не лежат в одной плоскости, поэтому равенство  $\boldsymbol{\theta} = \eta \mathbf{a} + \lambda \mathbf{b} + \mu \mathbf{c} = 0$  приводит к тому, что  $\eta = \lambda = \mu = 0$ . Итак, уравнения (П1.5) действительно линейно независимы.

Если учесть дополнительное соотношение  $\alpha_{ij}^2 + \beta_{ij}^2 = \gamma_{ij}^2 + \delta_{ij}^2 = 1$ , то уравнения (П1.5) имеют два решения, различающиеся только знаком. Выбирая одно из них, находим (с точностью до знака) составляющие векторов  $\mathbf{u}_{ij}$  относительно систем координат  $(p_i, q_i)$  и  $(p_j, q_j)$ . Далее можно определить расстояния:

$$\begin{aligned} d_1 &= \|\mathbf{u}_{12} - \mathbf{u}_{13}\|; \\ d_2 &= \|\mathbf{u}_{12} - \mathbf{u}_{23}\|; \\ d_3 &= \|\mathbf{u}_{13} - \mathbf{u}_{23}\|. \end{aligned} \quad (\text{П1.7})$$

Рассмотрим теперь треугольник со сторонами  $d_1, d_2$  и  $d_3$ . Если существует решение задачи реконструкции, то имеется, по крайней мере, один такой треугольник. Однако он может вырождаться, если хотя бы одно из расстояний оказывается равным нулю. В невырожденном случае этот треугольник единственный, и



все его вершины лежат на единичном расстоянии от начала координат\*. Таким образом, три вершины и начало координат определяют две симметричные друг другу пирамиды. Для каждой пирамиды известны проекции точек  $A$ ,  $B$  и  $C$  на три плоскости, которые определяют единственную трехмерную конфигурацию.

*Вырожденный случай.* Если одно из расстояний  $d_i=0$  ( $i=1, 2, 3$ ), то все они равны нулю и пирамида вырождается в линию. В противном случае совпали бы две плоскости  $\pi_i$  ( $i=1, 2, 3$ ), что противоречило бы предположению об их различии. Прежде чем рассматривать вырожденный случай, докажем две леммы.

Пусть  $p_1=(x_1, y_1)$ ;  $p_2=(x_2, y_2)$  и  $q_1=(x'_1, y'_1)$ ;  $q_2=(x'_2, y'_2)$  есть четыре точки, лежащие на некоторой плоскости  $(x, y)$  такие, что точки  $q_1$  и  $q_2$  не являются отражением точек  $p_1$  и  $p_2$  относительно оси  $y$ , и пусть линии  $p_1q_1$  и  $p_2q_2$  (которые будем называть траекториями) параллельны оси  $x$ . Повернем теперь пару точек  $q_1$  и  $q_2$  на угол  $\alpha$  относительно начала координат  $O$ . В результате они переходят в точки  $q'_1=(u_1, v_1)$  и  $q'_2=(u_2, v_2)$ . Считается, что точки  $O$ ,  $p_1$  и  $p_2$  не лежат на одной прямой.

**Лемма 1.** *Существует точно один угол  $\alpha>0$ , такой, что линии  $p_1q'_1$  и  $p_2q'_2$  параллельны.*

**Доказательство.** Для  $i=1, 2$  справедливо

$$\begin{aligned} u_i &= x'_i \cos \alpha - y'_i \sin \alpha; \\ v_i &= x'_i \sin \alpha + y'_i \cos \alpha. \end{aligned} \quad (\text{П1.8})$$

Если в результате поворота рассматриваемые линии остаются параллельными, то их наклоны одинаковы:

$$(y_1 - v_1)/(x_1 - u_1) = (y_2 - v_2)/(x_2 - u_2). \quad (\text{П1.9})$$

(Возможна ситуация, когда эти линии параллельны оси  $y$ , тогда знаменатели дробей оказываются равными нулю. Как будет показано, при этом решение также единственное.)

Используя выражения (П1.8) для  $u$  и  $v$  и учитывая при этом, что  $y'_i=y_i$  для  $i=1, 2$ , получаем

$$\frac{y_1 - x'_1 \sin \alpha - y_1 \cos \alpha}{x_1 - x'_1 \cos \alpha + y_1 \sin \alpha} = \frac{y_2 - x'_2 \sin \alpha - y_2 \cos \alpha}{x_2 - x'_2 \cos \alpha + y_2 \sin \alpha}. \quad (\text{П1.10})$$

Это соотношение сводится к уравнению

$$a \sin \alpha + b \cos \alpha = b, \quad (\text{П1.11})$$

где  $a=x_1x'_2-x'_1x_2$ ;  $b=x_1y_2+x'_1y_2-y_1x'_2$ .

Для заданных  $a$  и  $b$  это уравнение имеет точно одно решение для  $\alpha$ , которое определяется

$$\begin{aligned} \sin \alpha &= 2ab/(a^2 + b^2); \\ \cos \alpha &= (b^2 - a^2)/(a^2 + b^2). \end{aligned} \quad (\text{П1.12})$$

Только в случае, когда одновременно  $a$  и  $b$  равны нулю, отсутствует единственное решение для (П1.12). При этом из (П1.11) следуют два уравнения относительно  $x'_1$  и  $x'_2$ . Если эти уравнения независимы, то их решение  $x'_1=-x_1$  и  $x'_2=-x_2$  противоречит условию леммы о том, что точки  $q_1$  и  $q_2$  не являются отражением точек  $p_1$  и  $p_2$  относительно оси  $y$ . При этом справедливо соотношение  $x_1/x_2=y_1/y_2=x'_1/x'_2$ , которое противоречит предположению о коллинеарности. Если существует поворот  $\beta$ , обращающий в нуль знаменатели в (П1.9), то  $\beta$  все же является решением, и единственным, уравнения (П1.11)  $\blacklozenge$ .

**Лемма 2.** *Если проекции двух объектов  $O$  и  $O'$  на фронтальную плоскость совпадают и если это совпадение сохраняется при повороте обоих объектов на один и тот же угол  $\psi$  ( $\psi$  меньше  $180^\circ$ ) относительно вертикальной оси, то объекты  $O$  и  $O'$  конгруэнтны.* Доказательство этого утверждения очевидно и опускается.

\* Предполагается, что точка пересечения плоскостей  $\pi_1$ ,  $\pi_2$  и  $\pi_3$  (когда она существует) совпадает с началом фиксированной системы координат. — *Прим. пер.*

Наша цель состоит в том, чтобы установить единственность интерпретации для конфигурации  $(O, A, B, C)$ , вращающейся относительно некоторой фиксированной оси. Пусть ось вращения совпадает с осью  $z$  системы координат с началом в точке  $O$  и пусть плоскость  $(y, z)$  представляет собой плоскость изображения. Если интерпретация рассматриваемого объекта  $\Omega = (O, A, B, C)$  не единственная, то существует объект  $\Omega' = (O, A', B', C')$ , такой, что его повороты отличаются от поворотов объекта  $\Omega$ , но три проекции объектов  $\Omega$  и  $\Omega'$  совпадают. (Согласно лемме 2, если повороты одинаковые, то эти объекты конгруэнтны.)

Предположим, что между первым и вторым изображениями объект  $\Omega$  поворачивается на угол  $\alpha_1 \neq 0$ , а объект  $\Omega'$  — на угол  $\beta_1$ . Предположим также, что между вторым и третьим изображениями объект  $\Omega$  поворачивается на угол  $\alpha_2 \neq 0$ , а объект  $\Omega'$  — на угол  $\beta_2$ . При поворотах проекции объектов  $\Omega$  и  $\Omega'$  на плоскость изображения  $(y, z)$  совпадают. Пусть  $p_1 = (x_1, y_1)$  и  $p_2 = (x_2, y_2)$  — проекции соответственно точек  $A$  и  $B$  на плоскость  $(x, y)$ , а  $q_1 = (x'_1, y'_1)$  и  $q_2 = (x'_2, y'_2)$  — проекции соответственно точек  $A'$  и  $B'$  на ту же плоскость. Без потери общности можно считать, что точки  $p_1, p_2$  и точки  $q_1, q_2$  удовлетворяют требованиям леммы 1 (поскольку такие две пары должны существовать, если не для проекций точек  $A, B$  и  $A', B'$ , то для проекций точек  $A, C$  и  $A', C'$ ).

**Утверждение:**  $\alpha_2 = \beta_2$ .

**Доказательство:** между первым и вторым изображениями объект  $\Omega$  (т. е.  $p_1$  и  $p_2$ ) поворачивается на угол  $\alpha_1$ , объект  $\Omega'$  (т. е.  $q_1$  и  $q_2$ ) поворачивается на угол  $\beta_1$ , а траектории  $p_1q_1$  и  $p_2q_2$  остаются параллельными оси  $x$ . Если бы объект  $\Omega$  был неподвижным, а объект  $\Omega'$  был бы повернут на угол  $\beta_1 - \alpha_1$ , то эти траектории остались бы параллельными друг другу (хотя и не параллельными оси  $x$ ). В соответствии с леммой 1 существует единственный угол, для которого это произойдет. Назовем этот угол  $\psi$ , тогда  $\beta_1 - \alpha_1 = \psi$ . Между первым и третьим изображениями объект  $\Omega$  поворачивается на угол  $\alpha_1 + \alpha_2$ , объект  $\Omega'$  — на угол  $\beta_1 + \beta_2$ , а в результате траектории опять оказываются параллельными. Аналогично из существования единственного угла  $\psi$  следует, что  $(\beta_1 + \beta_2) - (\alpha_1 + \alpha_2) = \psi$ . Но поскольку  $\beta_1 - \alpha_1 = \psi$ , то  $\beta_2 = \alpha_2$ . Таким образом, при совместном вращении обоих объектов между вторым и третьим изображениями сохраняется совпадение их проекций. В соответствии с леммой 2 эти объекты конгруэнтны  $\blacklozenge$ .

Вышеприведенное доказательство предлагает метод для реального вычисления трехмерной структуры по трем ортогональным проекциям. Эти вычисления должны основываться на измеримых величинах, представляющих собой двухмерные координаты четырех точек на трех изображениях, на которых заданы соответственно системы координат  $(p_i, q_i)$  для  $i=1, 2, 3$ . В уравнениях (П1.5) эти величины используются для определения образующих пирамиду единичных векторов  $u_{ij}$ . Если пирамида не вырождается в линию, то для последующего определения трехмерной конфигурации уже достаточно двух изображений. Положение точки в пространстве можно найти как пересечение перпендикуляров, восстановленных из проекций этой точки на соответствующие две плоскости. Рассмотренное доказательство не дает способа определения структуры в вырожденном случае, но ее можно непосредственно определить, используя очевидные тригонометрические соотношения [Ульман, 1977а, приложение 2]. Существуют и другие алгоритмы, в большей степени удовлетворяющие требованию биологической осуществимости (см. разд. 3.3), но они не будут здесь обсуждаться.

## Приложение 2. Восстановление структуры по перспективным проекциям

Проблема, которая здесь рассматривается, состоит в вычислении структуры и движения по перспективным проекциям. Даются два *моментальных снимка*, соответствующих перспективным изображениям движущегося сохраняющего форму объекта, который содержит некоторое число идентифицируемых характерных точек. Предполагается, что уже известно соответствие между этими двумя картинками.

Следует заметить, что структуру и сдвиг можно определить по таким изображениям только с точностью до масштабного множителя. Нельзя сказать определенно, является ли объект *маленьким и расположенным близко* или *большим и расположенным далеко*. Более точно, пусть  $O$  — некоторый объект (набор трехмерных координат  $x_i, y_i, z_i$ ), который подвергается сохраняющему форму преобразованию  $R+T$ , где  $R$  — поворот, а  $T$  — сдвиг. Перспективные проекции этого движущегося объекта не отличимы от объекта  $cO$  (набора трехмерных координат  $cx_i, cy_i, cz_i$ ), который также подвергается сохраняющему форму преобразованию  $R+cT$ .

Сначала рассматривается случай, когда движение наблюдаемого объекта ограничивается комбинациями поворотов относительно вертикальной оси и произвольных сдвигов, затем — случай чистого сдвига. Новизна предлагаемого подхода заключается в том, что указанные случаи сводятся к решению единственного уравнения, заданного в явной форме. Произвольное движение обсуждаться не будет. Для дальнейшего ознакомления см. работу [Ульман, 1977а, приложение 3].

**Случай поворота относительно вертикальной оси.** Структура и движение восстанавливаются, если в каждый момент работать с тремя точками. Пусть эти три точки имеют координаты  $(x_1, y_1, z_1)$ ,  $(x_2, y_2, z_2)$  и  $(x_3, y_3, z_3)$  и подвергаются общему для них сохраняющему форму преобразованию  $R+T$ , где  $R$  — поворот на  $\theta$  градусов относительно вертикальной оси  $z$ , а  $T$  — вектор сдвига  $(\Delta x, \Delta y, \Delta z)$ . Координаты  $i$ -й точки, получающейся из  $i$ -й точки в результате указанного преобразования, имеют вид

$$\begin{aligned}x'_i &= x_i \cos \theta - y_i \sin \theta + \Delta x; \\y'_i &= x_i \sin \theta + y_i \cos \theta + \Delta y; \\z'_i &= z_i + \Delta z.\end{aligned}\tag{П2.1}$$

Хотя координаты  $(x_i, y_i, z_i)$  не восстанавливаются по изображению, но все же можно определить  $\varphi_i = x_i/y_i$  (смещение) и  $\eta_i = z_i/y_i$  (высота). В результате сохраняющего форму преобразования смещение и высота становятся равными соответственно

$$\begin{aligned}\varphi'_i &= x'_i/y'_i = (x_i \cos \theta - y_i \sin \theta + \Delta x)/(x_i \sin \theta + y_i \cos \theta + \Delta y); \\ \eta'_i &= z'_i/y'_i = (z_i + \Delta z)/(x_i \sin \theta + y_i \cos \theta + \Delta y).\end{aligned}\tag{П2.2}$$

Если подставить в (П2.2) вместо  $x_i$  и  $z_i$  соответственно  $x_i = \varphi_i y_i$  и  $z_i = \eta_i y_i$ , то останутся семь неизвестных параметров:  $y_1, y_2, y_3$  (определяющие структуру) и  $\theta, \Delta x, \Delta y$  и  $\Delta z$  (определяющие движение). Поскольку интерпретация осуществляется с точностью до масштабного множителя, реально имеется только шесть неизвестных, которые нужно определить, и шесть (нелинейных) уравнений: три для смещения и три для высоты. Громоздкие выкладки сводят эти шесть уравнений в единственное уравнение только относительно  $\theta$ :

$$\begin{aligned} & \alpha_3 \beta_1 \beta_2 (\varphi'_1 - \varphi'_2) + \alpha_2 \beta_1 \beta_3 (\varphi'_3 - \varphi'_1) + \alpha_1 \beta_2 \beta_3 (\varphi'_2 - \varphi'_3) = \\ & = \alpha_1 \alpha_2 \beta_3 (\eta'_2 - \eta'_1) + \alpha_1 \alpha_3 \beta_2 (\eta'_1 - \eta'_3) + \alpha_2 \alpha_3 \beta_1 (\eta'_3 - \eta'_2),\end{aligned}\tag{П2.3}$$

где  $\alpha_i = \varphi'_i \varphi_i \sin \theta + \varphi'_i \cos \theta + \varphi_i \cos \theta + \sin \theta$ ;  $\beta_i = \eta'_i \varphi_i \sin \theta + \eta'_i \cos \theta \eta_i$ .

После большого числа утомительных операций умножения, приведения подобных членов и упрощения уравнение (П2.3) принимает вид

$$A \sin^2 \theta + B \cos^2 \theta + C \cos \theta \sin \theta + D \sin \theta + E \cos \theta = 0.\tag{П2.4}$$

Уравнение (П2.4) будет называться *полярным уравнением*.

Далее приводятся коэффициенты этого уравнения —  $A, B, C, D$  и  $E$ . Выражения для них довольно сложные, однако они полностью выражаются через измеряемые значения  $\varphi_i$  и  $\eta_i$ . В этих выражениях можно сделать некоторые упрощения, если выбрать соответствующим образом систему координат, например если ось  $y$  выбрать так, что  $\varphi_1 = \eta_1 = 0$ ;  $A$  — коэффициент при  $\sin^2 \theta$ :

$$\begin{aligned} & ((1 + \varphi_1 \varphi'_1) \eta'_2 \eta_3 (1 + \varphi_2 \varphi'_3)) - \\ & ((1 + \varphi_1 \varphi'_1) \eta_2 \eta'_3 (1 + \varphi'_2 \varphi_3)) + \\ & ((1 + \varphi_2 \varphi'_2) \eta_1 \eta'_3 (1 + \varphi'_1 \varphi_3)) - \end{aligned}$$

$$\begin{aligned}
 & ((1 + \varphi_2 \varphi'_2) \eta'_1 \eta_3 (1 + \varphi_1 \varphi'_3)) + \\
 & ((1 + \varphi_3 \varphi'_3) \eta'_1 \eta_2 (1 + \varphi_1 \varphi'_2)) - \\
 & ((1 + \varphi_3 \varphi'_3) \eta_1 \eta'_2 (1 + \varphi'_1 \varphi_2));
 \end{aligned}$$

$B$  — коэффициент при  $\cos^2 \theta$ :

$$\begin{aligned}
 & (\eta'_2 \eta_3 (\varphi'_1 - \varphi_1) (\varphi'_3 - \varphi_2)) + \\
 & (\eta_2 \eta'_3 (\varphi'_1 - \varphi_1) (\varphi_3 - \varphi'_2)) + \\
 & (\eta_1 \eta'_3 (\varphi'_2 - \varphi_2) (\varphi'_1 - \varphi_3)) + \\
 & (\eta'_1 \eta_3 (\varphi'_2 - \varphi_2) (\varphi_1 - \varphi'_3)) + \\
 & (\eta'_1 \eta_2 (\varphi'_3 - \varphi_3) (\varphi'_2 - \varphi_1)) + \\
 & (\eta_1 \eta'_2 (\varphi'_3 - \varphi_3) (\varphi_2 - \varphi'_1));
 \end{aligned}$$

$C$  — коэффициент при  $\cos \theta \sin \theta$ :

$$\begin{aligned}
 & (\eta'_2 \varphi'_1 \varphi_2 \varphi'_3 (\eta_3 - \eta_1)) + \\
 & (\eta'_1 \varphi_1 \varphi'_2 \varphi'_3 (\eta_3 - \eta_3)) + \\
 & (\eta'_3 \varphi'_1 \varphi'_2 \varphi_3 (\eta_1 - \eta_2)) + \\
 & (\eta_1 \varphi'_1 \varphi_2 \varphi_3 (\eta'_2 - \eta'_3)) + \\
 & (\eta_2 \varphi_1 \varphi'_2 \varphi_3 (\eta'_3 - \eta'_1)) + \\
 & (\eta_3 \varphi_1 \varphi_2 \varphi'_3 (\eta'_1 - \eta'_2)) + \\
 & (\eta_1 \varphi'_2 (\varphi_2 - \varphi'_1) (1 + \varphi_3 \varphi'_3)) + \\
 & (\eta_1 \eta_2 (\varphi'_2 - \varphi_1) (1 + \varphi_3 \varphi'_3)) + \\
 & (\eta'_1 \eta_3 (\varphi_1 - \varphi'_3) (1 + \varphi_2 \varphi'_2)) + \\
 & (\eta_1 \eta'_3 (\varphi'_1 - \varphi_3) (1 + \varphi_2 \varphi'_2)) + \\
 & (\eta_2 \eta'_3 (\varphi_3 - \varphi'_2) (1 + \varphi_1 \varphi'_1)) + \\
 & (\eta'_2 \eta_3 (\varphi'_3 - \varphi_2) (1 + \varphi_1 \varphi'_1)) + \\
 & ((\varphi_3 - \varphi'_3) (\eta_1 \eta'_2 - \eta'_1 \eta_2)) + \\
 & ((\varphi_2 - \varphi'_2) (\eta_3 \eta'_1 - \eta'_3 \eta_1)) + \\
 & ((\varphi_1 - \varphi'_1) (\eta_2 \eta'_3 - \eta'_2 \eta_3));
 \end{aligned}$$

$D$  — коэффициент при  $\sin \theta$ :

$$\begin{aligned}
 & ((1 + \varphi_1 \varphi'_1) ((\eta'_2 \eta'_3 \varphi_2) + (\eta_2 \eta_3 \varphi'_2)) - \\
 & ((1 + \varphi_1 \varphi'_1) ((\eta'_2 \eta'_3 \varphi_3) + (\eta_2 \eta_3 \varphi'_3)) + \\
 & ((1 + \varphi_2 \varphi'_2) ((\eta'_1 \eta'_3 \varphi_3) + (\eta_1 \eta_3 \varphi'_3)) - \\
 & ((1 + \varphi_2 \varphi'_2) ((\eta'_1 \eta'_3 \varphi_1) + (\eta_1 \eta_3 \varphi'_1)) + \\
 & ((1 + \varphi_3 \varphi'_3) ((\eta'_1 \eta'_2 \varphi_1) + (\eta_1 \eta_2 \varphi'_1)) - \\
 & ((1 + \varphi_3 \varphi'_3) ((\eta'_1 \eta'_2 \varphi_2) + (\eta_1 \eta_2 \varphi'_2));
 \end{aligned}$$

$E$  — коэффициент при  $\cos \theta$ :

$$\begin{aligned}
 & ((\varphi_1 - \varphi'_1) \eta_2 \eta_3 (\varphi'_3 - \varphi'_2)) + \\
 & ((\varphi_1 - \varphi'_1) \eta'_3 \eta'_3 (\varphi_3 - \varphi_2)) + \\
 & ((\varphi_2 - \varphi'_2) \eta_1 \eta_3 (\varphi'_1 - \varphi'_3)) + \\
 & ((\varphi_2 - \varphi'_2) \eta'_1 \eta'_3 (\varphi_1 - \varphi_3)) + \\
 & ((\varphi_3 - \varphi'_3) \eta_1 \eta_2 (\varphi'_2 - \varphi'_1)) + \\
 & ((\varphi_3 - \varphi'_3) \eta'_1 \eta'_2 (\varphi_2 - \varphi_1)).
 \end{aligned}$$

**Число решений полярного уравнения.** Полученное полярное уравнение можно преобразовать в уравнение четвертой степени относительно  $\sin \theta$ , отсюда следует, что для  $\sin \theta$  имеется не более четырех решений. Два изображения трех точек не определяют единственным образом решение для  $\theta$  и, следовательно, структуру.

Один из возможных способов получения единственного решения заключается в использовании полярного уравнения для более сложных ситуаций, нескольких триплетов точек, и затем в исследовании множеств найденных решений для различных триплетов. Если пересечение этих множеств содержит единственное решение, то оно выбирается как искомое значение  $\theta$ . Машинная реализация позволила испытать полярное уравнение на большом числе примеров движущихся объектов и показала, что правильный ответ обычно можно найти по не менее чем двум изображениям четырех точек. Ниже приводится иллюстрация того, как определяется единственное решение.

Были даны два изображения объекта, содержащего пять точек. Начальные координаты этих точек составляли:  $P_1=(50, 200, 25)$ ;  $P_2=(20, 300, 60)$ ;  $P_3=(-10, 100, 15)$ ;  $P_4=(70, 160, -10)$ ;  $P_5=(-30, 230, -40)$ .

Преобразование заключалось в повороте  $R$  ( $\theta=30^\circ$ ) и в сдвиге  $T$  ( $\Delta x=15$ ;  $\Delta y=30$ ;  $\Delta z=20$ ).

Решения полярного уравнения для точек  $P_1$ ,  $P_2$  и  $P_3$  были равны  $11,5$  и  $30^\circ$ , а для точек  $P_4$ ,  $P_2$  и  $P_5$  —  $2,3$ ;  $20,1$  и  $30^\circ$ . Истинный поворот уже можно было определить, так как только решение, равное  $30^\circ$ , появилось в обоих множествах решений. Применение полярного уравнения к точкам  $P_1$ ,  $P_2$  и  $P_5$  дало результат, равный  $-13,4$  и  $30^\circ$ , подтверждающий предыдущее решение.

**Восстановление структуры и сдвигов.** После того как определен угол поворота  $\theta$ , можно также вычислить структуру и сдвиги объекта. Напомним, что в случае перспективной проекции решение получается с точностью до масштабного множителя. Примем здесь в качестве единицы измерения расстояния для структуры и сдвига  $y_1$  (т. е. глубину первой точки), поэтому вычисленное  $y_2$  на самом деле является отношением между  $y_2$  и  $y_1$ . То же справедливо и для других параметров.

*Вычисление сдвига.* Сдвиг в глубину  $\Delta y$  можно найти по любой из точек  $P_i$  по формуле

$$\Delta y = (\alpha_i \beta_1 - \alpha_1 \beta_i) / [\alpha_i (\eta'_i - \eta'_1) - \beta_i (\varphi'_i - \varphi'_1)], \quad (\text{П2.5})$$

откуда  $\Delta x = \alpha_1 + \varphi'_1 \Delta y$ ;  $\Delta z = \beta_1 + \eta'_1 \Delta y$ .

*Вычисление структуры.* Координаты глубины для точек  $P_i$  получаются с помощью уравнения

$$y_i = [\beta_1 + \Delta y (\eta'_i - \eta'_1)] / \beta_i, \quad (\text{П2.6})$$

откуда можно найти  $x_i$  и  $z_i$ , зная выражения для  $\varphi_i$  и  $\eta_i$ .

**Применение полярного уравнения.** Предположим, что делаются аэрофотоснимки некоторой местности. Если оптическая ось фотокамеры принадлежит некоторой плоскости (например, камера помещена в самолет, осуществляющий горизонтальный полет по прямой, или в спутник, находящийся на круговой орбите), то движение местности относительно камеры подпадает под категорию  $z$ -вращения и сдвигов. Следовательно, с помощью полярного уравнения наблюдаемая местность может быть отображена (с точностью до масштабного коэффициента) по двум снимкам, *даже если отсутствуют данные о траектории камеры или о ее положении относительно этой траектории*. Интересно отметить, что получаемая в результате информация почти столь же полна, как и при стереоскопическом отображении, хотя ни положение, ни ориентация камеры не известны (они определяются в процессе вычислений).

**Случай чистого сдвига.** Отсутствие поворота при движении объекта приводит с учетом перспективы к особенно простым вычислениям. Случай движения в глубину не является исключением: как любой другой чистый сдвиг, его легко вычислить. Если отсутствует поворот ( $\theta=0$ ), то полярное уравнение принимает вид  $B+E=0$ . Это простое соотношение можно применить в качестве *теста* на чистый сдвиг, полезность которого зависит от того, можно ли это уравнение (назовем его *тестом на сдвиг*) надежно использовать для выявления чистого сдвига. Полярное уравнение гарантирует, что в случае, когда некоторое множе-

ство точек действительно участвует в чистом сдвиге, каждый триплет будет удовлетворять тесту на сдвиг. Обратное, однако, не обязательно верно. Точки могут удовлетворять тесту, не будучи подвергнутыми чистому сдвигу. Применение теста к упомянутым примерам показало, что он имеет высокую надежность даже для нескольких точек. (Действительно, необходимое число точек зависит от точности вычислений — в использовавшейся машинной реализации обычно было достаточно четырех точек.)

*Вычисление структуры.* Для  $\theta=0$  коэффициенты  $\alpha_i$  и  $\beta_i$  принимают простую форму:  $\alpha_i = \varphi'_i - \varphi_i$  и  $\beta_i = \eta'_i - \eta_i$ . Если изменение  $\varphi_i$  между двумя картинками обозначить как  $\delta\varphi_i = \varphi'_i - \varphi_i$  и если аналогичным способом обозначить изменение  $\eta_i$ , наконец, если принять, что  $\Delta\varphi_i = \varphi'_i - \varphi'_1$  и  $\Delta\eta_i = \eta'_i - \eta'_1$ , то в таких обозначениях структура задается уравнением

$$y_i = (\delta\eta_1 \Delta\varphi_2 - \delta\varphi_1 \Delta\eta_2) / (\delta\eta_2 \Delta\varphi_2 - \delta\varphi_2 \Delta\eta_2). \quad (\text{П2.7})$$

Заметим, что структура определяется только относительными величинами. Действительные положения точек несущественны, так как решение основывается только на расстояниях между точками и на их смещениях между двумя моментальными снимками.

## Примечания

### К введению.

<sup>1</sup> Точное определение «идеализированных условий» не будет дано, поскольку это потребует рассуждений, выходящих за рамки введения. Вместо этого приведем иллюстративный пример. Теорию арифметических операций можно рассматривать как теорию, лежащую в основе электронного калькулятора, хотя ей в некоторых случаях и не удастся предсказать его действительное поведение. Когда такое расхождение существует, его можно приписать дефекту калькулятора. Другими словами, работа калькулятора считается успешной до той степени, до какой она удовлетворяет теории арифметики, и поэтому арифметику можно считать теорией калькулятора (или теорией его «идеализированной работы»).

<sup>2</sup> Утверждение, что кажущееся и действительное движения существенно не отличаются друг от друга, является следствием таких факторов: а) при определенных условиях кажущееся движение нельзя субъективно отличить от действительного; б) некоторые исследования (например, [Клатворти и Фрисби, 1973]) позволяют сделать вывод, что один и тот же механизм зрительной системы человека обеспечивает оба типа видимого движения; в) тот факт, что восприятие кажущегося движения было обнаружено у различных животных, наводит на мысль о его универсальной важности для восприятия естественного движения. Оно, вероятно, является врожденным как у людей [Таубер и Коффлер, 1966], так и у низших животных [Рок, Таубер и Хеллер, 1964]; г) прерывающееся движение может быть действительным в том смысле, что его можно встретить в естественных ситуациях; вот некоторые из возможных источников такой дискретности: саккадические движения глаза, мигания глаз и, наконец, расположение в глазе перед слоем рецепторов кровяных сосудов [Грегори и Гомбрих, 1973, с. 73].

Ответ на вопрос, имеет ли смысл различие между непрерывным и дискретным движениями, зависит от способа обнаружения и анализа движений. Рассмотрим следующую гипотетическую схему определения движения. Предположим, что детектор движения некоторого элемента состоит из двух отдельных детекторов, каждый из которых настроен на обнаружение этого элемента в определенном месте. Такие гипотетические детекторы движения по своей природе не могут находиться на слишком низком уровне. Одна из причин этого состоит в том, что кажущееся движение можно выявить внутри глаза, например когда при движении глаза два стимула оказываются в одном и том же месте сетчатки [Рок и Эбенхолтц, 1962]. Юлеш и Пэйн (1968) показали, что можно также установить движение между бинокулярно-воспринятыми линиями скачков интенсивности на стереограммах из случайных точек. В такой схеме не имеет значения, перемещается ли рассматриваемый элемент от одного места к другому

непрерывно или дискретно за один шаг. Одно из различий, которое действительно существует между этими двумя случаями, заключается в том, что промежуточные детекторы движения будут возбуждены только непрерывным движением. Важность явления кажущегося движения (в такой схеме обнаружения движения) состоит в том, что восприятие движения может быть вызвано без участия промежуточных детекторов. Это вряд ли удивительно, если учесть указанное выше соображение (г). В зависимости от реального механизма обнаружения движения, существующего в зрительной системе, возможно, единственным различием между восприятным кажущегося и действительного движений является то, что в первом случае используется только часть того механизма, который работает в последнем случае. Конечно, отсюда не следует, что указанные два вида движений ничем не отличаются друг от друга на любом уровне. Например, кажущееся движение можно вызвать дихоптически, т. е. осуществляя первое предъявление одному глазу, а второе — другому. Никто не осмелится утверждать, что такое движение должно иметь, например, те же самые зрительные характеристики, что и действительное непрерывное движение. Я уверен, что обнаруженные к настоящему времени различия между этими видами движения [Колерс, 1964; Кауфман и другие, 1971] имеют такую природу.

<sup>3</sup> Как было указано Канеманом и Вольманом (1970), преимущество в выборе времени предъявления, составляющего около 120 мс, заключается в том, что уменьшается влияние времени предъявления и тем самым МСИ становится существенно зависящей от времени переменной.

<sup>4</sup> На таком расстоянии информация, получаемая от процесса аккомодации к глубине, становится бесполезной [Пауэр и Дэй, 1973].

#### К гл. 1.

<sup>1.1</sup> Процесс вычитания подобен дискретной перекрестной корреляции, в которой умножение заменяется вычитанием, а именно вычисляется сумма вида  $\sum |A_i - B_i|$ , где  $A$  и  $B$  являются сравниваемыми массивами. Как и перекрестная корреляция, процесс вычитания может быть *глобальным*, когда сравниваются массивы целиком, или *локальным*, когда в массиве  $A$  ищется наиболее подходящее сопоставление для подобласти  $B$ .

<sup>1.2</sup> Простейший случай, в котором форма изображения не меняется, соответствует вращательному движению объекта относительно глаза, как, например, движение Луны вокруг Земли.

<sup>1.3</sup> См. определения, сделанные в примечании <sup>1.1</sup>. Графики для случая локальной обработки были получены путем ограничения подобласти  $B$  точками 160 и 200. Скачок интенсивности на картине  $B$  находится в точке 180.

<sup>1.4</sup> Дополнительные соображения в пользу этого утверждения следуют из различных психологических исследований, которые свидетельствуют о том, что в зрительной системе человека обработка формы и движения осуществляется раздельно [Толхерст, 1973; Фрисби и Клэвортс, 1974; Секьюлер и Левинсон, 1977].

#### К гл. 2.

<sup>2.1</sup> Я бы хотел определить общую точку зрения на то, что процесс нахождения соответствия основан на сопоставлении первичных опорных элементов, а не на какой-либо другой из схем, предложенных в прошлом. Я не считаю адекватными схемы, которые основаны на таких законах, как законы простейшего дуги и полной ассимиляции ([Шиллер, 1933] в [Холвэнд, 1935]), и я согласен с критикой Колерсом (1972) векторной модели, предложенной Брауном и Вотом (1937).

<sup>2.2</sup> Мишо, например, изучал так называемый эффект тоннеля в рамках своего исследования по феноменальной тождественности. В этих экспериментах движущийся объект на части своей траектории исчезает, как если бы он вошел в невидимый тоннель, а затем появляется (на другом конце тоннеля). Изучался вопрос, при каких условиях объект сохранял бы идентичность самому себе. Аналогично в некоторых других экспериментах, описанных Пиаже (1970), идентичность объектов выводилась с помощью рассуждений и сравнений, основанных на прежнем опыте, а не с помощью непосредственного восприятия.

<sup>2.3</sup> Термин «поле» здесь следует понимать с точки зрения психологии (см., например, [Коффка, 1935]).

2.4 Примеры, описанные здесь, взяты из работы Тернуса (1926). Необходимо заметить, что виды соответствия, которые он описал, не всегда воспринимаются реально. Различные виды соответствия и некоторые факторы, влияющие на них, обсуждаются в разд. 2.4.

2.5 Джоунс и Брунер (см. [Вернон, 1966, с. 244]) в явной форме выразили это предположение.

2.6 В случае эквидистантных ситуаций одни наблюдатели отдают некоторое предпочтение правой стороне (ср. [Колерс, 1972, с. 65]), а другие — левой. В настоящем изложении этот эффект можно считать пренебрежимо малым.

2.7 Расстояния между отрезками на предъявляемом изображении должны быть достаточно большими для того, чтобы соответственные устанавливалось между самими этими отрезками как опорными элементами, а не между их составными частями (см. разд. 2.4.2).

2.8 Одной из проблем, связанных с простейшим методом конкурирующих движений, является выполнение условий покрытия, рассмотренных в гл. 3. Процесс нахождения соответствия отдает предпочтение сопоставлениям, в которых ни один из элементов не остается без пары. Существуют различные способы усовершенствования этого простейшего метода конкурирующих движений, но они не будут здесь подробно рассматриваться.

2.9 Основные требования заключаются в том, чтобы функция сродства была монотонно возрастающей и конкурирующие взаимодействия были монотонно убывающими в зависимости от степени соответствия (см. разд. 2.3).

2.10 Другой способ определения периферийного прироста, периферийного выравнивания и выравнивания за счет МСИ состоит в том, чтобы считать сродство также функцией положения в зрительном поле и МСИ. Определение в терминах эффектов обращает особое внимание на одинаковое влияние, оказываемое смещением от центра и МСИ на сродство между парами опорных символов.

2.11 Влияние цвета, относительно которого имеются противоречивые сведения, не исследовалось. Холвэнд (1935) привел некоторые факты относительно того, что цвет играет незначительную роль в определении соответствия, тогда как в работе [Фернбергер, 1934] имеются обратные данные (хотя они могут быть артефактом, связанным с влиянием интенсивности света).

2.12 Фрисби и Юлеш (1976) наблюдали при стереоскопическом восприятии похожее соответствие между концами отрезков линий. Возможно, в совокупности опорных элементов включаются концы отрезков линий и, может быть, углы.

2.13 Все расстояния в этом эксперименте превышали 15' углового зрения для того, чтобы исключить возможное влияние процесса с малой зоной действия, упомянутого в гл. 1.

2.14 Понятие степени соответствия содержит только внешнее сходство с использовавшимися ранее понятиями притяжения и способности к сцеплению [Коффка, 1935; Браун и Вот, 1937], несостоятельность которых показал Колерс (1972).

2.15 Простые обобщения этой конфигурации конкурирующих движений показывают, что не только более слабая, но также и более сильная связь подвергается влиянию в результате конкуренции.

2.16 Небольшие одинаковые пятна ведут себя, если исследовать сродство, подобно точкам, рассмотренным в разд. 2.2. Для изучения зависимости сродства между пятнами от таких параметров, как размер и ориентация, потребуются дальнейшие исследования.

2.17 Перцептуальные сопоставления, описанные здесь, находятся в соответствии с результатами Колерса (1972); примеры 1, 3 и 4 соответствуют примерам 13, 15 и 16 в этой работе.

2.18 Тернус (1974) использовал изображение, состоящее из трех точек, в отличие от рассмотренной в этом примере двухточечной конфигурации.

2.19 Гештальтпсихологи обычно уделяли особое внимание только когерентному способу установления соответствия; ссылку на феномен двух способов установления соответствия (без учета влияния МСИ) можно найти в [Фернбергер, 1934].

2.20 Эттивн (1974) высказал идею, что за возникновение способа установления соответствия по связанному элементу ответственна линия, «мысленно» про-



водная между этими элементами. Такая линия, являющаяся частным случаем группирования, рассмотренного в подразд. 2.4.2, сама по себе не может объяснить *переход* от одного способа к другому. Более того, способ установления соответствия по связанному элементу имеет место для элементов, подобных параллельным линиям, существование между которыми «мысленно» проведенной линии маловероятно. Более подробно конфигурация Тернуса обсуждается в гл. 3.

2.21 Следует заметить, что для саккадического подавления, вероятно, достаточно только движений глаза без сдвига изображения на сетчатке.

### К гл. 3.

3.1 Можно было бы выбрать другую общую стратегию, минимизирующую ожидаемую ошибку. Она не обязательно максимизирует вероятность правильного ответа.

3.2 В случае, когда группирование предшествует сопоставлению, соседние элементы, конечно, будут двигаться вместе, но здесь это несущественно. Нужно также заметить, что, хотя положительные взаимодействия, похоже, не являются необходимыми, все же их нельзя полностью отбросить. В соответствии с гл. 2 полезно не принимать во внимание дополнительные взаимодействия, если только их необходимость серьезно не обоснована.

3.3 В зрительном отделе коры головного мозга большинство тангенциальных связей довольно коротко (см. [Сентаготаи, 1973]). Тангенциальные расстояния между большинством клеток ограничены несколькими сотнями микрон, только для слоя I и корзинчатых клеток в других слоях тангенциальные расстояния больше и достигают нескольких миллиметров.

3.4 Доказательство этого утверждения здесь не приводится. Подробнее изложение см. в [Ульман, 1978с].

3.5 От начальных кандидатов следует потребовать, чтобы они все располагались недалеко друг от друга. Нарушение этого требования — одна из возможных причин того, что элемент может остаться без пары. Если он находится слишком далеко от других элементов, то скорее всего этот элемент не имеет начальных кандидатов для сопоставления.

3.6 Такая пороговая операция ускорит сходимость, но полученное решение, вероятно, будет отличаться от идеального минимального отображения. Однако представляется более разумным предпочесть быстрое приближение к минимальному отображению точным, но гораздо более медленным вычислениям.

3.7 Используется следующее определение сходимости: для любого  $\epsilon > 0$  существуют такие  $\rho$  и  $\tau$ , что итеративная процедура стремится к седловой точке для  $t < \tau$  и останется в  $\epsilon$ -окрестности седловой точки для  $t > \tau$  для любой длины шага, меньшей  $\rho$ . Более подробно см. [Эрроу и другие, 1958, гл. 10].

3.8 Соотношение  $d = v \cdot \text{МСИ}$  справедливо только приблизительно, поскольку время предъявления каждой картинке также необходимо принимать во внимание [Канеман и Волман, 1970].

3.9 Например, распределение соответствий не меняется, когда пары точек из работы [Колерс, 1972, массивы 22, 23] заменяются короткими отрезками линий. Однако оно меняется, когда добавляются фоновые точки, нарушающие группирование. Поэтому, похоже, соответствие в этих примерах иногда устанавливается между опорными символами групп, а не между отдельными точками. Возможность влияния пересекающихся путей на процесс заполнения (см. гл. 5) связана с правилом непересечения траекторий.

3.10 Похожая способность обнаруживать сдвиг существует для статических изображений. Однако она, вероятно, основывается на других принципах [Стивенс, 1978].

3.11 Если поворот достаточно мал, то он, конечно, также будет обнаружен. Заметим, что лемма об обнаружении сдвига не будет применима к схеме минимального отображения, когда смещения слишком большие, поскольку это отображение не будет больше локальным покрытием.

3.12 В перспективной проекции измеряется угловая скорость  $\omega = v/r$ . Если не используется информация о глубине, то всем точкам можно приписать одно  $n$  то же  $r$ . Следовательно, выражение (3.23) можно было бы выразить через  $\omega$  вместо  $v$ . Необходимо также заметить, что предположения, использовавшиеся

для получения  $p(v)$ , не являются решающими; они используются в основном для получения оценки вида  $p(v)$ .

<sup>3.13</sup> Варьируя  $v$  изменением МСИ между картинками, необходимо соблюдать предосторожность по отношению к влиянию времени предъявления. Если время предъявления менее 120 мс, то его увеличение приводит к тем же результатам, что и увеличение МСИ, и поэтому  $v_i$  не определяется просто как  $d_i/\text{МСИ}$ .

<sup>3.14</sup> Примечательно, что минимальное отображение было разработано в условиях предъявления двух картинок с относительно небольшим числом опорных символов и в условиях значительного времени предъявления. Обобщение такого эксперимента, описанного в разд. 3.8, представляет собой минимальное обобщение в том смысле, что не учитываются дополнительные взаимодействия. В качестве примера возможного дополнительного взаимодействия можно рассматривать «коллинеарное усиление» между связями, которое могло бы усилить (локально) линейные траектории. Другой возможностью является ослабление требования покрытия по мере приближения к периферии зрительного поля, где элементы имеют большую вероятность появляться и исчезать.

<sup>3.15</sup> Этот термин был предложен А. Уиткином (Массачусетский технологический институт).

#### К гл. 4.

<sup>4.1</sup> Поскольку список литературы, относящейся к этому вопросу, весьма длинный, он приводится здесь отдельно: [Майлс, 1931; Гибсон, 1950; Уоллок и О'Коннэлл, 1953; Уоллок, О'Коннэлл и Нейссер, 1953; Гибсон и Гибсон, 1957; Гибсон, 1957; Фиандт и Гибсон, 1959; Гибсон, Гибсон, Смит и Флок, 1959; Грин, 1961; Браунштейн, 1962; Йоханссон, 1964; Йоханссон и Янссон, 1963, 1973; Хэй, 1966; Борьессон и Хофстен, 1972, 1973; Йоханссон, 1973, 1974, 1975; Эрикссон, 1974; Бэдлер, 1975].

<sup>4.2</sup> Демонстрация конвейерной ленты служит и другой цели. Далее в этом разделе вводится понятие *неявных предположений*. Эта демонстрация показывает, что при определении конфигурации конвейерной ленты не используется неявное предположение о постоянной скорости (ср. [Хофстен, 1974]).

<sup>4.3</sup> Гибсон и другие обратили внимание на то, что при одновременном предъявлении двух движущихся плоскостей, неизменно плоскость с большей скоростью не воспринималась как расположенная ближе плоскости с меньшей скоростью.

<sup>4.4</sup> Частота, возможно, не является единственным уместным соображением. Например, наплавляющий стимул можно интерпретировать либо как увеличение объекта, либо как его приближение. Возможно, *безопаснее* всегда интерпретировать такую ситуацию как приближающуюся угрозу.

<sup>4.5</sup> Понятия ограничений и предположений можно часто использовать одинаковым образом. Например, процесс интерпретации может быть основан на гипотезе, что утверждение  $P$  справедливо. Эквивалентно он может быть ограничен с целью получения решений, которые удовлетворяют  $P$ . Термину «ограничение» иногда отдаются предположения с тем, чтобы явным образом описать способ применения этих ограничений. Ограничение скорее предполагает использование фильтра, который в универсальной неизменной форме допускает только некоторые из возможных интерпретаций, в отличие от гипотезы, которая могла бы быть предложена, оценена и пересмотрена.

<sup>4.6</sup> Хомский (1965) внес различие между компетентностью и действительной работой (см. также [Марр, 1977a] и примечание 1 во введении, касающееся понятия «теория»).

<sup>4.7</sup> Алгоритм параллельного поиска был разработан как часть исследования по использованию параллельных вычислительных машин для обработки зрительной информации в институте перспективных методов вычисления, Саннывейл, Калифорния.

<sup>4.8</sup> Основное требование неявного предположения заключается в его выполнении в течение большей части времени наблюдения. Если некоторое утверждение почти всегда оказывается ложным, то полезное неявное предположение можно получить с помощью отрицания этого утверждения. Наихудшим критерием, таким образом, является такой, который выполняется почти случайно. Оценка поворота с помощью двигательного параллакса, которая оказывается

верной лишь для времени, незначительно меньшего 50% всего времени наблюдения, не может поэтому иметь большую пользу.

<sup>4.9</sup> Для более подробного ознакомления с «отменой» поворота см. приложение 2 в работе [Ульман, 1977а]. Необходимо заметить, что для описанной схемы требуется знание относительной глубины точек *A* и *B* (см. рис. 4.8), которые находятся далеко друг от друга, что делает указанную схему более подверженной ошибкам.

#### **К гл. 5.**

<sup>5.1</sup> В случае, когда различие между этими двумя интерпретациями мало, потребуются более точные эксперименты для определения, берется ли в некотором смысле усредненная интерпретация.

<sup>5.2</sup> Тот факт, что Сл.Д. схема в благоприятных условиях при восприятии доминирует, затрагивает интересный вопрос о силе восприятия, которая может служить мерой достоверности, связанной с рассматриваемым перцептом. Такую меру можно было бы использовать для различных зрительных перцептов, но в настоящей работе это изучаться не будет.

<sup>5.3</sup> Один из таких эффектов представляет собой уже упомянутую тенденцию интерпретировать стимулы, соответствующие максимальному удлинению, как лежащие на фронтальной плоскости. Другой эффект связан с возможным влиянием памяти. При неоднозначных условиях определение направления вращения зависит от непосредственной предыстории движения.

## СПИСОК ЛИТЕРАТУРЫ

- Ames, A. 1951. Visual perception and the rotating trapezoid window. *Psychological Monographs*, Vol. 65 (7), Whole No. 324.
- Anstis, S. M. 1970. Phi movement as a subtraction process. *Vision Research* 10, 1411-1430.
- Anstis, S. M. and Rogers, B. J. 1975. Illusory reversal of visual depth and movement during changes of contrast. *Vision Research*, 15, 957-961.
- Arnheim, R. 1974. Art and visual perception Berkely and Los Angeles: University of California Press.
- Arrow, K. J., Hurwicz, L., and Uzawa, H. 1958. *Studies in Linear and Nonlinear Programming* Stanford: Stanford University Press.
- Attneave, F. and Frost, R. 1969. The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics*, (6B), 391-396.
- Attneave, F. and Block, G. 1973. Apparent motion in tridimensional space. *Perception and Psychophysics*, Vol. 13, 2, 301-307.
- Attneave, F. 1972. Representation of physical space. In: A. W. Melton and E. Martin (Ed.) *Coding Processes in Human Memory* Washington: Winston and Sons.
- Attneave, F. 1974. Apparent movement and the what-where connection. *Psychologia*, 17, 108-120.
- Badler, N. I. 1975. Temporal scene analysis: Conceptual descriptions of object movements. *Technical Report 80*, Department of Computer Science, Univ. of Toronto.
- Bell, H. H. and Lappin, J. S. 1973. Sufficient conditions for the discrimination of motion. *Perception and Psychophysics* Vol. 14, 1, 45-50.
- Borjesson, Erik and von Hofsten, Claes. 1972. Spatial determinants of depth perception in two-dot motion patterns. *Perception and Psychophysics* Vol. 11, 4 263-268.
- Borjesson, Erik and von Hofsten, Claes. 1973. Visual perception of motion in depth: Application of a vector model to three dot motion patterns. *Perception and Psychophysics* Vol. 13, 2, 169-179.

Bower, T. G. R. 1966. The visual world of the infants. *Scientific American*, 225, 80-92.

Bower, T. G. R. 1966. Slant perception and shape constancy in infants. *Science*, 151, 832-834.

Bower, T. G. R. 1971. Slant perception and shape constancy in infants. *Scientific American*, 225, 30-38.

Bower, T. G. R. 1974. *Development in Infancy*. San Francisco: W. H. Freeman.

Braddick, O. 1974. A short-range process in apparent motion. *Vision Research*, 14, 519-527.

Braunstein, M. L. 1962. Depth perception in rotation dot patterns: effects of numerosity and perspective. *Journal of Experimental Psychology*, Vol. 64, 4, 415-420.

Brown, J. F. and Voth, A. C. 1937. The path of seen movement as a function of the vector field. *American Journal of Psychology*, Vol. 49, 543-563.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: The M.I.T. Press.

Clatworthy, J. L. and Frisby, J. P. 1973. Real and apparent movement: evidence for a unitary mechanism. *Perception*, 2(2) 161-164.

Cohen, B. L. and Salapatek, P. 1975. *Infant Perception From Sensation to Cognition*. New York, San Francisco, London: Academic Press.

Corbin, H. H. 1942. The perception of grouping and apparent motion in visual space. *Archives of Psych. Whole No. 273*.

Cornsweet, T. 1970. *Visual Perception*. New York: Academic Press.

Cross, J. and Cross, J. 1969. The misperception of rotary motion. *Perception and Psychophysics*, Vol. 5, (2), 94-96.

Davson, H. 1969. *The Eye*. New York: Academic press.

Day, R. H. and McKenzie, B. E. 1973. Perceptual shape constancy in early infancy. *Perception*, 2, 315-320.

Ditchburn, R. W. 1973. *Eye Movements and Visual Perception*. Oxford: Clarendon Press.

Eden, M. 1962. A three-dimensional optical illusion. *Quarterly Progress Report No. 64* MIT R.L.E. 267-274.

Ellis, W. D. 1967. *A Source book of Gestalt Psychology* New York: Humanities Press.

Epstein, w. Jansson, G. and Johansson, G. 1968. Perceived angle of oscillatory motion. *Perception and Psychophysics Vol. 3, 1A*, 12-16.

Eriksson, E. S. 1974. A theory of veridical space perception. *Scandinavian Journal of Psychology, 15*, 225-235.

Fernberger, S. W. 1934. New phenomena of apparent visual movement. *American Journal of Psychology, 46*, 309-314.

Feynman, R. 1965. *The Character of Physical Law* Cambridge, Mass.: The M.I.T. Press.

Fieandt, von Kai and Gibson, J. J. 1959. The sensitivity of the eye to two kinds of continuous transformations of a shadow-pattern. *Journal of Experimental Psychology, 57*, 344-347.

Fremlin, D. 1977. Personal Communication.

Frisby, J. P. 1972. The effect of stimulus orientation on the Phi phenomenon. *Vision Research, 12*, 1145-1166.

Frisby, J. P. and Clatworthy, J. L. 1974. Evidence for separate movement and form channels in the human visual system. *Perception, Vol. 3*, 87-96.

Frisby, J. P. and Julesz, B. 1976. The effect of length difference between corresponding lines on stereopsis from single and multi-line stimuli. *Vision Research, 16*, 83-87

Garfinkel, R. S. and Nemhauser, G. L. 1972. *Integer Programming*. New York: Wiley & Sons.

Gibson, E. J., Gibson, J. J., Smith, O. W. and Flock, H. 1959. Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology, Vol. 8, 1*, 40-51.

Gibson, E. J. 1969. *Principles of Perceptual Learning and Development*. New York: Appleton-Century-Crofts.

Gibson, J. J. 1950. *The perception of the visual world* Boston: Houghton Mifflin and Co.

Gibson, J. J. 1957. Optical motions and transformations as stimuli for visual perception. *Psychological review*, Vol. 64 (5), 288-295.

Gibson, J. J. and Gibson, E. J. 1957. Continuous perspective transformation and the perception of rigid motion. *Journal of Experimental Psychology*, Vol. 54, 2, 129-138.

Gibson, J. J. 1959. Perception as a function of stimulation. In: S Kotch (Ed.) *Psychology A Study of a Science* New York: McGraw-Hill.

Gibson, J. J. 1965. Research on the visual perception of motion and change. In: Spigel, I. [1965] 125-146.

Gibson, J. J. 1966. *The senses considered as a perceptual system* Boston: Houghton Mifflin.

Gibson, J. J. 1968. What gives rise to the perception of motion? *Psychological Review*, Vol. 75, 4, 335-346.

Graham, C. H. 1963. On some aspects of real and apparent visual movement. *Journal of the Optical Society of America* Vol. 53 9, 1019-1025.

Graham, C. H. (ed.) 1965. *Vision and Visual perception* New York: Wiley.

Green, B. F. 1961. Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology*, Vol. 62, 3, 272-282.

Gregory, R. L. 1970 *The Intelligent Eye*. N.Y., St. Louis, San Francisco:

Gregory, R. L. and Gombrich, E. H. (ed.) 1973. *Illusions in Nature and Art* Duckworth.

Hay, C. J. 1966. Optical motions and space perception an extension of Gibson's analysis. *Psychological Review*, 73, 550-565.

Helmholtz, H. L. F. von. 1910. *Treatise on physiological optics*. Translated by J. P. Southall, 1925, N.Y. Dover Publications.

Hershberger, W. A. 1967. Comment on apparent reversal (oscillation) of rotary motion in depth. *Psychological Review*, 74 235-238.

Hershberger, W. A. & Urban, D. 1970. Depth perception from motion parallax in one dimensional polar projection: projection versus viewing distance. *Journal of Experimental Psychology*, 86, 133-136.

- Hershberger, W. A. & Urban, D. 1970. Three motion parallax cues in one-dimensional polar projection of rotation in depth. *Journal of Experimental Psychology*, 86, 380-383.
- Hershberger, W. A. & Starzec, J. J. 1974. Motion parallax cues in one dimensional polar and parallel projections: Differential velocity and acceleration/displacement change. *Journal of Experimental Psychology*, 103 (4), 717-723.
- von Hofsten, C. 1974. Proximal velocity change as a determinant of space perception. *Perception and Psychophysics*, Vol. 15, 3, 488-494.
- Holvand, C. I. H. 1935. Apparent movement. *Psychological Bulletin.*, 32, 755-778.
- Jansson, G. and Johansson, G. 1973. Visual perception of bending motion. *Perception*, Vol. 2, 321-326.
- Johansson, G. 1964. Perception of motion and changing form. *Scandinavian Journal of Psychology*, 5, 181-208.
- Johansson, G. and Jansson, G. 1968. Perceived rotary motion from changes in a straight line. *Perception and Psychophysics*, Vol. 4, 3, 165-170.
- Johansson, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, Vol. 14, 2, 201-211.
- Johansson, G. 1974. Visual perception of rotary motion as transformation of conic sections -- a contribution to the theory of visual space perception. *Psychologia*, 17, 226-237.
- Johansson, G. 1975. Visual Motion Perception. *Scientific American*, Vol. 232, No. 6, 76-88.
- Jones, E. E. and Bruner, J. S. 1954. Expectancy in apparent visual movement. *British Journal of Psychology*, 45, 157-165.
- Julesz, B. and Payne, R. A. 1968. Differences between monocular and binocular stroboscopic movement perception. *Vision Research*, 8, 433-444.
- Julesz, B. 1971. *Foundation of Cyclopean Perception*. Chicago: The University of Chicago Press.
- Kabrisky, M. 1966. *A Proposed Model for Visual Information Processing in the Human Brain* Urbana and London: The University of Illinois Press.



Kahneman, D. and Wolman, R. 1970. Stroboscopic motion: effects of duration and interval. *Perception and Psychophysics*, Vol. 8, 3, 161-164.

Kaufman, L. Cyrulnik, I., Kaplowitz, J., Melnick, G. and Stoff, D. 1971. The complementarity of apparent and real motion. *Psychologische Forschung*, 34, 343-348.

Koffka, K. 1935. *Principles of Gestalt Psychology* (280-306). New York: Harcourt, Brace and World.

Kolers, P. A. 1964. The illusion of movement. *Scientific American*, 211 (4), 98-106.

Kolers, P. A. 1966. An illusion that dissociates motion, object and meaning. *Quarterly Progress Report No. 82*, M.I.T R.L.E. 221-223.

Kolers, P. A. 1968. Some psychological aspects of Pattern Recognition. In: Kolers, P. A. and Edén, M. (ed.) 1968. *Recognizing Patterns*. Cambridge and London: The MIT Press.

Kolers, P. A. 1972. *Aspects of Motion Perception*. New York: Pergamon Press.

Korte, A. 1915. Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, 72, 193-296.

Kuhn, H. W. and Tucker, A. W. 1951. Nonlinear Programming. In J. Neyman (ed.) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* Berkeley and Los Angeles: University of California Press, 481-492.

Leese, J. A., Novak, C. S. and Taylor, V. R. 1970. The determination of cloud pattern motion from geosynchronous satellite image data. *Pattern Recognition*, 2, 279-292.

Lindsay, P. H. and Norman, D. A. 1972. *Human Information Processing* New York and London: Academic Press.

Mach, Ernst. 1959. *The Analysis of Sensations* New York: Dover Publications.

Mackay, D. M. 1961. The effects of non-redundant stimulation. *Nature (London)*, 192, 739-740.

Mackay, D. N. 1970. Elevation of visual threshold by displacement of retinal image. *Nature*, 225, 90-92.

- Marmolin, H. 1973. Visually perceived motion in depth resulting from proximal changes. *Perception and Psychophysics*, Vol. 14, 1, I: 133-142 II: 142-148.
- Marr, D. 1974. A note on the computation of binocular disparity in a symbolic, low-level visual processor. *M.I.T. A.I. Memo*, No. 326.
- Marr, D. 1976. Early Processing of visual information. *Philosophical Transactions of the Royal Society of London*, Vol. 275, 942, 483-534.
- Marr, D. 1977a. Artificial Intelligence -- A personal view. *Artificial Intelligence*, 9, 37-48.
- Marr, D. 1977b. Representing visual information. *M.I.T. A.I. Memo*, No. 415.
- Marr, D. and Poggio, T. 1976. Cooperative computation of stereo disparity.
- Marr, D. and Poggio, T. 1977. From understanding computation to understanding neural circuitry. *Neuroscience Research Program Bulletin*, 15(3), 470-488.
- Michotte, A. 1963. *The Perception of Causality*. New York: Basic books Inc. (original edition in 1946, translated by T. R. Miles).
- Miles, W. R. 1931. Movement interpretations of the silhouette of a revolving fan. *American Journal of Psychology*, 43, 392-505.
- Navon, D. 1976. Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology, Human Perception and Performance*, 2(1), 130-138.
- Neff, Walter S. 1936. A critical investigation of the visual apprehension of movement. *American Journal of Psychology* Vol. XLVIII, 1-42.
- Olson, R. K. 1974. Slant judgements from static and rotating trapezoids correspond to rules of perspective geometry. *Perception and Psychophysics*, 15(3), 509-516.
- Orlansky, J. 1940. The effect of similarity and difference in form on apparent visual movement. *Archives of Psychology*, 246.
- Ortega, J. M. and Rheinboldt, W. C. 1970. *Iterative Solution of Nonlinear Equations*. New York: Academic Press.

Pantle, A. J. 1973. Stroboscopic movement based upon global information in successively presented visual patterns. *Journal of the Optical Society of America*, 1280.

Pantle, A. J. and Picciano, L. 1976. A multistable display: Evidence for two separate motion systems in human vision. *Science*, Vol. 193, Aug. 6, 500-502.

Piaget, J. 1954. *The construction of reality in the child* New York: Basic Books.

Piaget, Jean. 1970. *The Childs Conception of Motion and Speed* Translated from French by G. E. T. Holloway and M. J. Mackenzie. New York: Basic Books.

Potter, J. 1974. The Extraction and utilization of motion in scene description. *Ph.D Thesis*, University of Wisconsin.

Power, R. P. and Day, R. H. 1973. Constancy and illusion of apparent direction of rotary motion in depth: Tests of a theory. *Perception and Psychophysics* Vol. 13, 2, 217-223.

Price K. 1975. A Comparison of human and computer vision. *ACM SIGART Newsletter* 50 5-10.

Ramachandran, V. S., Madhusudhan, V. R. and Vidyasagar, T. R. 1973. Apparent movement with subjective contours. *Vision Research*, Vol. 13, 1399-1401.

Ratliff, F. 1972. Contour and Contrast. *Scientific American*, 266, (6), 90-101.

Rock, I. and Ebenholtz, S. 1962. Stroboscopic movement based on change of phenomenal rather than retinal location. *American Journal of Psychology*, 72, 221-229.

Rock, I., Tauber, E. S. and Heller, D. P. 1964. Perception of stroboscopic movement: Evidence for its innate basis. *Science*, 147, 1050-1052.

Sekuler, R. and Levinson, E. 1977. The perception of moving targets. *Scientific American*, 236, (1), 60-73.

Shepard, R. N. and Judd, S. A. 1976. Perceptual illusion of rotation of three-dimensional objects. *Science*, Vol. 191, 952-954.

Smith, E. A. and Phillips, D. R. 1972. Automated cloud tracking using precisely aligned digital ATS pictures. *IEEE Trans. Computers*, 21, 715-729.

Sigman, E. and Rock, I. 1974. Stroboscopic movement based on perceptual intelligence. *Perception, Vol. 3*, 9-28.

Sperling, G. 1976. Movement perception in computer-driven visual displays. *Behavior research method and instrumentation, 8*, 144-151.

Spigel, I. (ed.) 1965. *Readings in the Study of Visually Perceived Motion* New York: Harper and Row.

Stevens, K. A. 1978. Computation of locally parallel structure. *Biological Cybernetics, 29*, 19-28

Szentagothai, J. 1973. Synaptology of the visual cortex. in: Jung, R. (ed.) 1973. *Handbook of Sensory Physiology. Vol. VII/3B*. Berlin Heidelberg New York: Springer-Verlag.

Tauber, E. S. and Koffler, S. 1966. Optomotor response in human infants to apparent motion: Evidence of innateness. *Science, 152*, 382-382.

Ternus, J. 1926. Experimentelle Untersuchung uber phanomenale Identitat. *Psychologische Forschung, 7*, 81-136. Translated in [Ellis 1967].

Toch, H. H. and Ittelson, W. H. 1956. The role of past experience in apparent motion: A revaluation. *British Journal of Psychology, 47*, 195-207.

Folhurst, D. J. 1973. Separate channels for the analysis of the shape and movement of a moving visual stimulus. *Jouranal of Physiology, 231*, 385-402.

Traub, J. F. 1964. *Iterative Methods for the Solution of Equations*. N.J.: Prentice-Hall.

Ullman, S. 1976a. On visual detection of light sources. *Biological Cybernetics 21*, 205-212.

Ullman, S. 1976b. Filling in the gaps: on the shape of subjective contours and a model for their generation. *Biological Cybernetics, 25*, 1-6.

Ullman, S. 1977a. The interpretation of visual motion. *M.I.T. Ph.D. Thesis, Department of Electrical Engineering and Computer Science*.

Ullman, S. 1977b. Transformability and object identity. *Perception and Psychophysics 22(4)*, 414-415.

Ullman, S. 1978a. Two dimensionality of the correspondence process in apparent motion. *Perception, 5* Forthcoming.

Ullman, S. 1978b. The interpretation of structure from motion. *Proceedings of the Royal Society of London*, Forthcoming.

Ullman, S. 1978c. Simple networks in the processing of visual information. *Forthcoming M.I.T. A.I. Memo*.

Vernon, M. D. (ed.) 1966. *Experiments in Visual Perception* Penguin Books.

Vernon M.-D. 1971. *The Psychology of Perception*. Pinguin Books.

Volkman, C. F. 1962. Vision during voluntary saccadic eye movements. *Journal of the Optical Society of America*, Vol. 52, 5, 571-578.

Wallach, H. and O'Connell, D. N. 1953. The Kinetic depth effect. *Journal of Experimental Psychology*, 45, 4, 205-217.

Wallach, H., O'Connell, D. N., and Neisser, U. 1953. The memory effect of visual perception of 3-d form. *Journal of Experimental Psychology*, 45, 360-368.

Wallach, H., Weisz, A. and Adams, P. A. 1956. Circles and derived figures in rotation. *American Journal of Psychology*, 69 48-59.

Wallach, H. 1959. The perception of motion. *Scientific American*, 201, 56-60.

Warren, H. W. 1977. Visual information for object identity in apparent motion. *Perception & Psychophysics*, 21, 264-268.

Weir, S. 1965. The perception of motion: actions, motives and feelings. *University of Edinburgh A.I report No. 13*.

Wertheimer, M. 1912. Experimentelle studien uber das sehen von bewegung. *Zeitschrift fur Psychologie*, 61, 161-265.

White, B. W. and Mueser G. E. 1960. Accuracy in reconstructing the arrangement of elements generating kinetic depth displays. *Journal of Experimental Psychology*, Vol. 60, 1, 1-11.

Wolferts, K. 1974. Special problems in interactive image processing for traffic analysis. *Second International Joint Conference on Pattern Recognition*, 1-2.

Zegers, R. T. 1964. The reversal illusion of the Ames trapezoid. *Transactions of the N.Y. Academy of Science*, 26, 377-400.

### **Список литературы, переведенной на русский язык**

- Арнхейм Р.** Искусство и визуальное восприятие. — М.: Прогресс, 1974.
- Грегори Р.** Разумный глаз. — М.: Мир, 1972.
- Линдсэй П., Норман Д.** Переработка информации у человека. — М.: Мир, 1974.
- Ортега Дж., Рейнболт У.** Итерационные методы решения нелинейных систем уравнений со многими переменными. — М.: Мир, 1975.
- Фейнман Р.** Характер физических законов. — М.: Мир, 1968.
- Хомский Н.** Аспекты теории синтаксиса. — М.: МГУ, 1972.
- Эрроу К., Гурвиц Л., Удзава Х.** Исследования по линейному и нелинейному программированию. — М.: ИЛ, 1962.

ШИМОН УЛЬМАН  
ПРИНЦИПЫ ВОСПРИЯТИЯ ПОДВИЖНЫХ ОБЪЕКТОВ

Редактор *Е. А. Засядько*  
Обложка художника *П. П. Рогачёва*  
Художественный редактор *Л. Н. Сильянов*  
Технические редакторы *Л. К. Грачева, Л. А. Горшкова*  
Корректор *Л. В. Алексеева*

ИБ № 269

---

Сдано в набор 15.02.83 г. Подписано в печать 12.05.83 г.  
Формат 60×90/16 Бумага тип. № 1 Гарнитура литературная Печать высокая  
Усл. печ. л. 10,5 Усл. кр.-отг. 10,875 Уч.-изд. л. 11,91 Тираж 4000 экз. Изд. № 20071  
Зак. № 33 Цена 1 р. 70 к.  
Издательство «Радио и связь». 101000 Москва, Почтамт, а/я 693

---

Типография издательства «Радио и связь» Госкомиздата СССР  
101000 Москва, ул. Кирова, д. 40