
MATHEMATICAL ELEMENTS FOR COMPUTER GRAPHICS

Second Edition

David F. Rogers

*Professor of Aerospace Engineering
United States Naval Academy, Annapolis, Md.*

J. Alan Adams

*Professor of Mechanical Engineering
United States Naval Academy, Annapolis, Md.*

McGraw-Hill Publishing Company

New York · St. Louis · San Francisco · Auckland · Bogotá · Caracas · Hamburg · Lisbon
London · Madrid · Mexico · Milan · Montreal · New Delhi · Oklahoma City · Paris
San Juan · São Paulo · Singapore · Sydney · Tokyo · Toronto

Д. Роджерс, Дж. Адамс

МАТЕМАТИЧЕСКИЕ ОСНОВЫ МАШИННОЙ ГРАФИКИ

Перевод со второго английского издания

П. А. Монахова, Г. В. Олохтоновой,
Д. В. Волкова

под редакцией

Ю. М. Баяковского, В. А. Галактионова,
В. В. Мартынюка



Москва «Мир» 2001

УДК 517.91
ББК 22.19
Р60

От редакторов перевода

Роджерс Д., Адамс Дж.

Р60 Математические основы машинной графики: Пер. с англ. —
М.: Мир, 2001. — 604 с., ил.

ISBN 5-03-002143-4

Полностью переработанное и дополненное второе издание книги известных американских специалистов (перевод первого издания вышел в издательстве «Машиностроение» в 1980 г.). Книгу отличает глубина и фундаментальность материала, четкий и лаконичный стиль изложения, удачное сочетание строгости подхода с практической направленностью. От читателя требуется знакомство с математикой в объеме вузовской программы и знание одного из языков программирования.

Для математиков-прикладников, для всех, кто занимается и интересуется машинной графикой, автоматизацией проектирования, для студентов вузов.

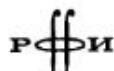
ББК 22.19

В 1990 г. в Далласе на конференции SIGGRAPH мы обсуждали с проф. Д. Роджерсом планы перевода и издания на русском языке данной книги. Уже через год перевод книги был подготовлен, но в силу известных обстоятельств выпуск ее задержался на много лет. И только теперь благодаря поддержке РФФИ издание книги стало возможным. Безусловно это знаковое событие. Оно свидетельствует о том, что происходит оживление и восстановление в российской науке, технике, промышленности.

Материал, изложенный в книге, не устарел. Впервые на русском языке появляется достаточно полное функциональное описание технических средств машинной графики, систематическое изложение двумерных и трехмерных преобразований, практическое руководство по геометрическому моделированию, включающее такие объекты, как кривые, поверхности, сплайны. Всего этого так не хватает нашим студентам, аспирантам, инженерам, научным работникам.

Мы искренне благодарны Д. Роджерсу за столь плодотворное сотрудничество и надеемся, что оно будет продолжено.

Ю. М. Баяковский, В. А. Галактионов



Издание осуществлено при поддержке Российского фонда
фундаментальных исследований по проекту № 99-01-14007

Редакция литературы по математическим наукам

Copyright © 1990, 1976 by Mc Graw-Hill, Inc.
© перевод на русский язык,
Монахов П. А., Олохтонова Г. В.,
Волков Д. В., 2001
© оформление, «Мир», 2001

ISBN 5-03-002143-4 (русск.)
ISBN 0-074053529-9 (англ.)

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Я с большой радостью узнал о выходе в свет в переводе на русский язык второго английского издания «Математических основ машинной графики». Эта книга составит прекрасный тандем вместе с ранее изданной в переводе на русский язык книгой «Алгоритмические основы машинной графики». Я надеюсь, что в совокупности эти две книги образуют фундаментальный материал по машинной графике и в какой-то степени они скомпенсируют существующую нехватку публикаций по этому предмету на русском языке.

Сегодня, даже в большей степени, чем в ту пору, когда была написана эта книга, машинная графика глубоко проникает в искусство, технику, рекламу, связь. Я упомянул лишь некоторые немногие области приложений, а список можно продолжить. В частности, следует отметить Интернет, который возник как средство быстрого и легкого обмена текстовой информацией между учеными, а в настоящее время он сильно зависит от технологии машинной графики.

Почти на каждом вэб-сайте для представления информации используется машинная графика. Многие спецэффекты в видео- и кинофильмах не обходятся без применения средств машинной графики.

В индустриальном мире машинная графика играет огромную роль, особенно в робототехнике, автоматизированном проектировании и производстве, в научной визуализации. В наши дни обычный человек, не задумываясь, может послать картинки своим друзьям и коллегам по электронной почте простым нажатием на кнопку «мышки». Отмечу и такой факт: объем бизнеса в машинной графике достиг 90 млрд. долларов.

За мои без малого три с половиной десятка лет деятельности в области машинной графики произошли поистине удивительные изменения. Как авиакосмический инженер я всегда восхищался прекрасными традициями замечательной российской науки и техники. Я полон ожиданий, что и в новых условиях становления и укрепления открытого общества русские ученые и инженеры внесут значительный выдающийся вклад в дальнейшее развитие машинной графики.

Дэвид Ф. Роджерс
Аннаполис, Мэриленд, США
1 декабря 2000 г.

ВВЕДЕНИЕ К ПЕРВОМУ ИЗДАНИЮ

Прошло уже более десяти лет с момента выхода в свет первого издания книги; за это время машинная графика привлекла к себе пристальное внимание множества специалистов из разных областей знания. В процентном отношении основное внимание в среде увеличивающейся плеяды профессионалов по машинной графике уделялось проблемам программирования, системного проектирования, аппаратуре и т. п. Именно это обстоятельство отмечал Сазерленд в своем предисловии к книге Принса «Interactive Graphics for Computer Aided Design», изданной в 1971 г. Такая ситуация наблюдается и сегодня. Однако я считаю, что уделяется недостаточно внимания именно прикладным проблемам, а это влечет за собой дефицит информации, которую можно было бы использовать при создании действительно работоспособных систем машинной графики. Основное принципиальное требование к разработке систем машинной графики состоит в необходимости глубокого понимания связи ее практических аспектов со спецификой конкретного приложения. Фундаментом большинства приложений машинной графики являются математические методы, особенно геометрия и способы преобразования. Тем не менее многие программисты и аналитики, работающие в области машинной графики, в своих работах обычно не рассматривают основы и не разбирают сложных математических задач. Более того, потенциал машинной графики часто остается полностью не реализованным именно вследствие неадекватного использования математического аппарата с его ограничениями и допущениями. Ценность книги профессоров Роджерса и Адамса состоит в ее практической ориентации. Четкий и лаконичный стиль изложения значительно облегчает задачу читателей по освоению материала. Книга является редким примером такой работы, которую многим практикам следует поместить на видное место в своей библиотеке в качестве незаменимого энциклопедического справочника. Она может быть также с успехом использована и в качестве основы учебного курса.

Я поздравляю авторов с созданием яркой и полезной книги «Математические основы машинной графики».

S. H. «Chas» Chasen
Lockheed Georgia Company, Inc.

ПРЕДИСЛОВИЕ

Предисловие

9

За четырнадцать лет, прошедших с момента издания первой редакции данной книги, машинная графика претерпела поистине грандиозные изменения и в настоящее время представляет собой важнейшую дисциплину для специалистов из многих областей знания, и в частности, для проектировщиков электронных устройств, инженеров, математиков, физиков и художников. В это время были в совершенстве разработаны прежние концепции машинной графики, а также предложены новые идеи. Однако основополагающие математические понятия остались неизменными. Глубокое овладение базовыми математическими концепциями до сих пор остается главным ключом к пониманию и дальнейшему развитию машинной графики. Предлагаемая книга содержит материал, необходимый для усвоения этих концепций.

Данная редакция книги не является результатом простого пересмотра первого издания, а представляет собой его полную переработку с включением множества новых иллюстраций и более подробных примеров. В книге содержатся формулировки проблем и концепций программных проектов. Алгоритмы, реализующие математические теории, приведены в книге в виде псевдокодов.

Глава 1 содержит новые данные о растровых устройствах и дисплеях, струйных и термических графопостроителях, цветных камерах. В главе приводится также перечень литературы по аппаратному обеспечению машинной графики.

В главе 2 содержится подробное описание двумерных преобразований, дополненное новыми разделами по преобразованиям сплошных тел, а также описание и геометрическая интерпретация однородных координат.

Глава 3 была значительно расширена новыми разделами и дополнена материалом по комбинированным преобразованиям, повороту вокруг произвольных осей, отражению относительно любой плоскости, косоугольным проекциям (проекциям кабине и кавалье), точкам бесконечности, фотографическим преобразованиям. В главе подробно обсуждаются проблемы создания стереоизображений, а также разбираются методы фиксированного и центрального проецирования.

Глава 4 в настоящей редакции включает развернутое описание коник и методов использования конических сечений.

Глава 5 была полностью переработана. Для достижения большей ясности и наглядности изложения в нее включено детальное рассмотрение вопросов параболического сглаживания, кривых Безье и В-сплайнов, а также обсуждение общенных параболических сплайнов. В главе исследуются согласования и деления В-сплайнов, а также периодические униформные В-сплайны. В главу включено развернутое обсуждение унифицированных и неунифицированных кривых В-сплайна (NURBS).

Глава 6 также претерпела значительные изменения. Теперь она включает изучение линейных и бикубических поверхностей, а также управляемых и динамических поверхностей. В главе нашли отражение сведения о поверхностях вращения, охвата и о квадратических поверхностях Безье и поверхностях В-сплайнов. Рассматриваются вопросы сшивки поверхностей В-сплайна, деления поверхности, а также вводятся понятия гауссовой кривизны и математически правильной поверхности. Достаточно подробно в главе разбираются нерациональные и рациональные, однородные и неоднородные поверхности В-сплайнов (NURBS).

Обширное приложение включает программы машинной графики, описание методов работы с матрицами, псевдокоды, формат файла задания поверхности В-сплайна, задачи, специфики программ и алгоритмы в виде псевдокодов.

Материал книги был тщательно подобран в расчете как на студентов, так и на аспирантов. Книга может быть использована в качестве учебного пособия для полугодового вводного курса машинной графики для студентов или аспирантов. (Двухсеместровый курс основывается на другой книге «Алгоритмические основы машинной графики», которая является естественным продолжением данной.) Именно таким образом авторы и использовали книгу «Математические основы машинной графики». Если в односеместровом курсе требуется изложение более широкого круга вопросов, то можно использовать обе книги. В этом случае предлагаются следующие темы: гл. 1 обеих книг, затем гл. 2 и 3 с отдельными разделами из гл. 4 (например от раздела 4-1 до 4-8) данной книги, далее отдельные разделы гл. 2 (например, от 2-1 до 2-5, 2-7 и от 2-15 до 2-19, 2-22, 2-23 и 2-28), гл. 3 (например, с разд. 3-1, 3-2, 3-4 по 3-6, 3-9, 3-11, 3-15 и 3-16), гл. 4 (например, разд. 4-1, часть разд. 4-2 с алгоритмом отбрасывания невидимых граней, разд. 4-3, 4-4, 4-7, 4-9, 4-11 и 4-13) и гл. 5 (разд. с 5-1 по 5-3, 5-5, 5-6 и 5-14) книги «Алгоритмические основы машинной графики». Д. Роджерс с успехом использовал данный материал для вводного и обзорных курсов. Надеемся, что наша книга пригодится также профессиональным программистам, инженерам и научным работникам. Благодаря детальному изложению, примерам из практики и многочисленным иллюстрациям, книга может быть полезна читателям разной квалификации для самостоятельного изучения. Для усвоения материала достаточно знания математики на уровне средней школы и знакомства с языком программирования высокого уровня.

Возможно, читателю будет интересно узнать, как создавалась книга. Она была набрана Нэнси А. Роджерс в системе обработки текстов TeX. При работе применялся комплекс шрифтов Computer Modern Roman. Все макроопределения, необходимые для согласования с требованиями издательства, были написаны Дэвидом Ф. Роджерсом. Для работы использовалось два компьютера: IBM AT и Zenith 386. Рукопись была введена непосредственно с бумажной копии. После ее редактирования на экране дисплея были получены с помощью лазерного принтера разрешением 300 точек на дюйм гранки, корректуры страниц и макеты. Окончательный вариант, пригодный для включения иллюстраций, был передан на фотонабор Американского Математического Общества в формате .dvi файлов системы TeX.

Книга не появилась бы на свет без содействия многих людей. Мы признательны студентам Центра прикладной физики Университета Джона Гопкинса,

которые просмотрели исходные варианты первых пяти глав книги. Многие высказанные студентами предложения и замечания оказались очень полезными.

Отдельно хочется поблагодарить Джона Дилла и Фреда Манчмейера, двух наиболее дорогих и близких нам коллег, прочитавших рукопись целиком с красным карандашом в руках. Многие высказанные ими предложения и замечания способствовали значительному улучшению книги. Особую признательность следует выразить нашей коллеге Линде Эдлам, которая не только прочитала всю рукопись, но и проверила все примеры. Огромное спасибо Стивену Д. Роджерсу, прочитавшему рукопись и проверившему примеры из первых пяти глав. Замечания Майка Джайгента об этих главах также были весьма ценными. Очень полезным был обзор работ по кривым и поверхностям Безье, а также кривым и поверхностям В-сплайна, проведенный Билом Гордоном.

Усилия Вирджинии Адамс по макетированию окончательного варианта книги также были очень ценными. Мы благодарны Барбаре Битон за ее неисчерпаемое терпение при консультировании нас по особенностям системы TeX. Специально отдадим должное Джастину Залмстре, чьи выполненные весьма кстати макроопределения для TeX'a существенно облегчили нашу задачу. Традиционно высокий профессионализм был продемонстрирован редактором издательства McGraw-Hill Джимом Брэдли. Наиболее сложные иллюстративные программы проверены Мэлом Хабером. Особую благодарность заслуживает Б. Д. Кларк, который вот уже около двадцати лет является нашим редактором в McGraw-Hill и был весьма внимателен к нашим иногда неортодоксальным идеям.

*Дэвид Ф. Роджерс, Дж. Алан Адамс,
Аннаполис, февраль 1989*

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Настоящее время характеризуется появлением новой быстро развивающейся области знания — «машинной графики», которая включает в себя как традиционные сферы знания, так и новые информационные технологии. Теперь компьютеры используются не только крупными фирмами и организациями, уже достигшими первых успехов в области программного и аппаратного обеспечения, но и широким кругом пользователей-непрофессионалов. Такое широкое внедрение компьютеров стало возможно благодаря недорогим графическим дисплеям, режиму разделения времени, а также успехам в области мини- и микрокомпьютерной технологии. Машинная графика является сегодня практическим, надежным, эффективным и широкодоступным инструментом.

Задача данной книги состоит в том, чтобы на основе универсального обобщения изложить математическую теорию основных методов машинной графики. Несмотря на использование в книге новых приемов представления материала, в ней нет совсем незнакомого читателю математического аппарата; в ней предпринята попытка собрать воедино всю информацию по машинной графике, разбросанную по разным техническим изданиям.

При отборе материала для книги мы уделяли особое внимание основным математическим по своей природе методам, а не методам процедурного характера. Именно по этой причине читатель обнаружит в данной книге прежде всего достаточно развернутое изложение способов поворота, преобразования, построения перспективных проекций, описание кривых и поверхностей, а не обсуждение задач отсечения или удаления невидимых линий и поверхностей. Знание математики в объеме первого курса колледжа является достаточной предпосылкой для освоения основной части книги.

После обсуждения в гл. 1 современной компьютерной технологии дается описание матричных операций над графическими элементами, заданными в однородных координатах. В последующих главах содержится описание современных способов представления точек, линий, кривых и поверхностей в цифровых компьютерах. Приводятся также соответствующие программные процедуры манипуляции и вывода графических изображений.

Вместе с математическими методами формирования аксонометрических и перспективных видов приводятся методы поворота, преобразования и масштабирования геометрических фигур. Процедуры определения кривых, заданных явно и в параметрической форме, описываются как для двумерного, так и для трехмерного пространства. Методы определения кривых предусматривают использование конических сечений, круговой интерполяции дугами, кубических сплайнов, параболического сглаживания, кривых Безье и кривых на основе В-сплайнов. В книгу включено также введение в математическое описание поверхностей.

Приложение содержит подпрограммы на языке Бейсик, реализующие алгоритмы большинства базовых модулей интерактивных пакетов машинной графики. Однако эти алгоритмы преднамеренно оканчиваются такими командами или подпрограммами реальной визуализации результатов, которые безусловно не являются стандартными для конкретного графического устройства. Несмотря на то, что в приложении А приводится некое предварительное описание графических примитивов и элементов, пользователь в общем случае считает необходимым ориентироваться на специфику конкретной физической конфигурации графических устройств.

Главные концепции данной книги были положены в основу вводного курса машинной графики для студентов высших учебных заведений научно-технической специализации. Курс хорошо подходит как для университетов, так и для технических школ. Книга полезна также и в качестве дополнительного материала для более углубленных учебных курсов по программированию или математике. Кроме этого, она может быть с успехом использована профессиональными программистами для индивидуальной подготовки. В конечном счете хорошо документированные программы могут быть полезны пользователям, заинтересованным в развитии графических возможностей своего компьютера.

БЛАГОДАРНОСТИ

Авторы глубоко признательны Военно-Морской Академии США за поддержку и внимание. Обеспеченная администрацией научная среда, предоставленное оборудование и, особенно, участие курсантов оказало весьма благоприятное воздействие на подготовку материала данной книги.

Ни одна книга не может быть написана без содействия многих людей, и мы хотели бы выразить здесь благодарность некоторым из них. Прежде всего Стиву Коонеу, просмотревшему всю рукопись и сделавшему множество ценных предложений, Ричарду Рейзенфельду, изучившему материал по кривым и поверхностям В-сплайна, профессору Пьеру Безье, который рецензировал материал по кривым и поверхностям Безье, а также Ивану Сазерленду, предложившему нам включить в гл. 3 обсуждение методов трехмерного преобразования. Особой благодарности заслуживают бывшие и нынешние члены группы САПР в Кэмбриджском университете. В первую очередь, это Робин Форест, Чарльз Ланг и Тони Нутбоул, проявившие глубокое понимание предмета машинной графики. Мы благодарим также Лоя Кноппа, написавшего оригинальную программу на Фортране для кривых В-сплайна.

Авторы хотели бы выразить признательность за содействие многим сотрудникам Evans & Sutherland Computer Corporation и особенно Джиму Каллану, который является автором документации, послужившей основой для многих идей по подготовке изображений, их представлению и интерактивному взаимодействию с ними. Особо отметим Ли Биллу, подготовившего все векторные рисунки.

Большая часть работы по художественному оформлению гл. 1 была проведена на рабочих станциях различных фирм-производителей. Отдельно выразим благодарность следующим корпорациям:

Рис. 1-3 Evans & Sutherland Computer Corporation

- Рис. 1-5 Adage Inc.
- Рис. 1-7 Adage Inc.
- Рис. 1-8 Vector General, Inc.
- Рис. 1-11 Xynetics, Inc.
- Рис. 1-12 CALCOMP, California Computers Products, Inc.
- Рис. 1-15 Gould, Inc.
- Рис. 1-16 Tektronix, Inc.
- Рис. 1-17 Evans & Sutherland Computer Corporation
- Рис. 1-18 CALCOMP, California Computers Products, Inc.

*Дэвид Ф. Роджерс,
Дж. Алан Адамс*

*Посвящается нашим женам:
Нэнси А. Роджерс и Вирджинии Ф. Адамс
и нашим семьям:
Стивену, Карен и Рэнсому Роджерс
и Линн, Дэвиду и Аллану Адамс*

Глава 1

ВВЕДЕНИЕ В МАШИННУЮ ГРАФИКУ

В настоящее время машинная графика представляет собой достаточно развитую отрасль знания, однако некоторые термины и определения по-прежнему трактуются очень широко. Например, такие понятия, как САПР, интерактивная графика, машинная графика, АСТПП, часто используются либо в разных значениях, либо так, что в значительной степени теряется их точный смысл. Наиболее типичным термином такого рода является «САПР», что означает любое применение компьютера для проектирования отдельных деталей, узлов или систем. Однако такое использование компьютера необязательно предусматривает включение машинной графики, ведь процесс проектирования может осуществляться на концептуальном уровне, на уровне деталировки либо ориентироваться на интерфейс с АСТПП.

В АСТПП компьютер используется для управления процессом промышленного производства. Для АСТПП требуется язык программирования, позволяющий задавать геометрию и траекторию перемещения обрабатывающего инструмента, например язык АРТ (язык программирования станков с ЧПУ), для генерирования команд управления инструментом станка. Контроллер станка, как правило, представляет собой микро- или миникомпьютер, для которого САПР может непосредственно формировать необходимые команды управления. Одновременно можно создавать массив данных в стандартном формате, например IGES (формат передачи графических данных). Для преобразования полученных данных в формат команд конкретного станка используются специальные программы. На рис. 1-1 показан типичный обрабатывающий центр с ЧПУ и контроллером.

Компьютер в машинной графике используется для описания изображений, их хранения, манипулирования ими, их вызова и визуализации, т. е. в основном для пассивных операций. Компьютер выводит на экран предварительно подготовленное изображение, и наблюдатель не может непосредственно управлять изображением, прежде чем оно появится на экране. В каждом конкретном случае изобра-

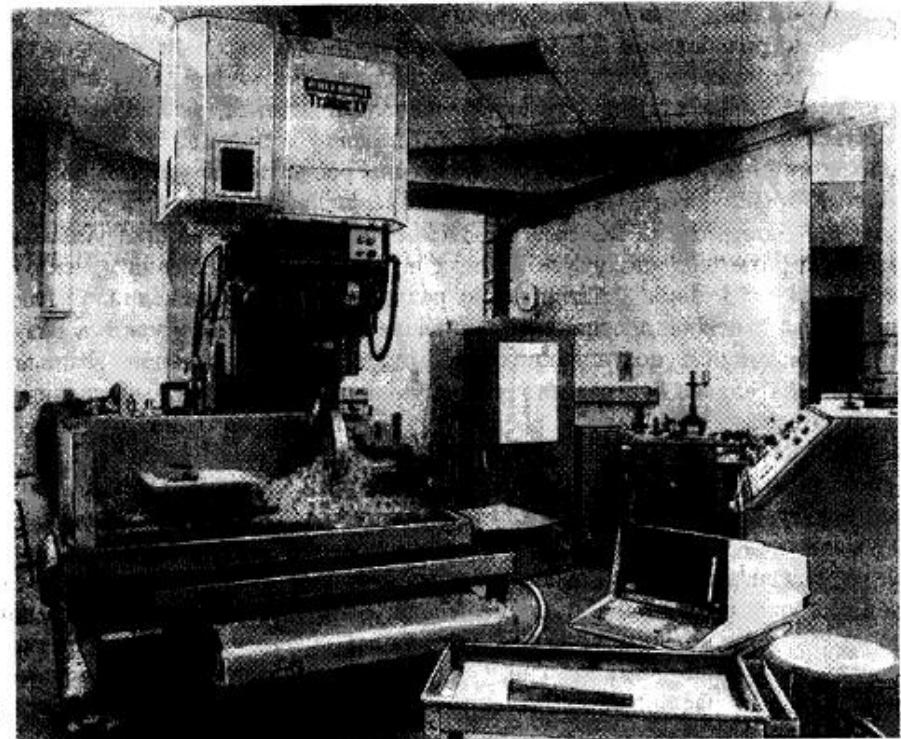


Рис. 1-1 Обрабатывающий центр с ЧПУ.

жение может быть либо таким простым, как, например, графики простых функций, либо столь сложным, как визуализация результатов моделирования процесса автоматического полета и посадки самолета или космического корабля.

В динамической интерактивной машинной графике (в дальнейшем просто интерактивной графике) компьютер используется для подготовки и визуализации графических данных. Однако интерактивная графика позволяет наблюдателю в реальном масштабе времени влиять на весь процесс представления изображения. Чтобы убедиться в серьезности ограничений, определяемых условиями работы в реальном времени, рассмотрим задачу вращения объемного изображения, состоящего из 1000 линий, со скоростью 15 град/с.

Как мы увидим в дальнейшем, картинка из 1000 линий в большинстве случаев представляется матрицей 1000×4 однородных координат конечных точек линий. Вращение, как правило, представляется путем умножения этой матрицы на матрицу преобразования размером 4×4 . Для проведения такого матричного умножения требуется выполнить 16 тыс. операций умножения, 12 тыс. операций сложения и 1 тыс. операций деления вещественных чисел. Если умножение осуществляется программно, то для того чтобы выполнить все операции вращения, может потребоваться значительное время. Обычный универсальный компьютер

с процессором вещественной арифметики, способным выполнять операции умножения с быстродействием 3.6 мкс, сложения — за 2.6 мкс и деления со скоростью 5.2 мкс, может выполнить такое матричное умножение за 0.1 с. Поскольку для того, чтобы картинка на экране начала двигаться, она должна перерисовываться не менее 30 раз в секунду, то становится очевидным, что в случае с данным примером картинка не сможет непрерывно перемещаться по экрану. Даже если предположить, что изображение обновляется только 15 раз в секунду, т. е. при повороте на каждый градус, то и в этом случае программная реализация не позволяет получить эффект непрерывного вращения и исключает возможность получения действительно реалистичных изображений. Существует множество способов достижения реального интерактивного режима визуализации изображения: использование более мощного дополнительного компьютера или более тщательное программирование, уменьшающее время выполнения матричного умножения. Однако в любом случае обязательно будет достигнут момент, когда всякое дальнейшее улучшение станет уже невозможным. Уменьшение сложности изображения также не даст нужного эффекта, так как при этом может оказаться неприемлемым качество картинки. Однако операция матричного умножения, необходимая для манипулирования изображением из нашего примера, для действительно очень сложных картинок может осуществляться на уровне микропрограмм или с помощью специализированных цифровых умножителей, применение которых всегда было наиболее эффективным и экономичным. На основе введенной терминологии в следующих разделах главы предлагается обзор машинной графики и обсуждаются различные типы современных графических устройств.

1-1 ОБЗОР МАШИННОЙ ГРАФИКИ

Машинная графика — это сложная и многогранная дисциплина. Поэтому для изучения ее необходимо разделить на более простые обозримые части, учитывая при этом, что конечным результатом машинной графики является изображение, которое, разумеется, можно использовать для самых различных целей: технического чертежа, деталировки в справочном руководстве, изображения архитектурной конструкции, как одного из возможных вариантов, так и эскиза проекта, рекламной иллюстрации либо отдельного кадра из мультфильма. В машинной графике фундаментальным связующим звеном является изображение, следовательно, важно разобрать следующие вопросы:

- представление изображений в машинной графике;
- подготовка изображений для вывода;
- визуализация предварительно подготовленных изображений;
- взаимодействие с изображением.

В дальнейшем под термином «изображение» мы будем понимать множество линий, точек, текстов и т. п., выводимых на графическое устройство.

1-2 ПРЕДСТАВЛЕНИЕ ИЗОБРАЖЕНИЙ

Многие алгоритмы интерпретируют изображения как набор многоугольников или ребер, причем каждый многоугольник или ребро в свою очередь может быть представлен с помощью точек (вершин). Таким образом, точки являются основными строительными блоками изображения. Не менее важны также и алгоритмы размещения таких точек. В качестве иллюстрации рассмотрим единичный квадрат в первом квадранте координатной плоскости. Этот квадрат может быть представлен своим четырьмя вершинами (см. рис. 1-2) $P_1(0, 0)$, $P_2(1, 0)$, $P_3(1, 1)$, $P_4(0, 1)$. Соответствующая алгоритмическая нотация может выглядеть следующим образом:

Последовательно соединить $P_1P_2P_3P_4P_1$.

Единичный квадрат можно также представить с помощью его четырех ребер $E_1 = P_1P_2$, $E_2 = P_2P_3$, $E_3 = P_3P_4$, $E_4 = P_4P_1$. Здесь алгоритмической нотацией служит следующая запись:

Последовательно изобразить $E_1 \ E_2 \ E_3 \ E_4$.

И наконец, для описания единичного квадрата в виде многоугольника можно использовать либо точки, либо ребра. Например, $S_1 = P_1P_2P_3P_4P_1$, или $S_1 = P_1P_4P_3P_2P_1$ или $S_1 = E_1E_2E_3E_4$. В зависимости от размерности координатного пространства основные строительные блоки (точки) можно представить либо как пары, либо как тройки чисел. Таким образом, координаты (x_1, y_1) или (x_1, y_1, z_1) определяли бы точку на плоскости или в пространстве. Две точки задавали бы отрезок или ребро, а совокупность из трех и более точек — многоугольник. Построение кривых осуществляется обычно путем их аппроксимации короткими отрезками. Символы представляются уже более сложным образом — они образуются из совокупности различных кривых или определяются матрицей точек. Однако в

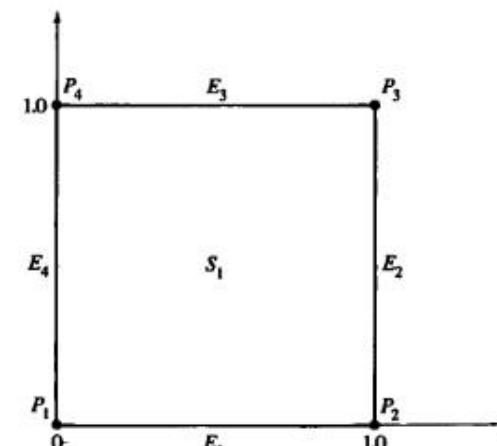


Рис. 1-2 Описание данных изображения.

основе представления символов по-прежнему лежат совокупности отрезков, точек и соответствующие им алгоритмы. Если не надо создавать специальные шрифты или проектировать графические устройства, то пользователю совсем не обязательно знать все детали описания символов текста, так как большинство графических устройств имеют встроенные аппаратные или программные генераторы символов.

1-3 ПОДГОТОВКА ИЗОБРАЖЕНИЙ ДЛЯ ВЫВОДА

В конечном счете изображение определяется точками и алгоритмом их вывода. Прежде чем информация об изображении будет использована для формирования картинки, она хранится в файле, называемом базой данных. Для представления очень сложных изображений нужны сложные базы данных, для которых предлагаются использование сложных алгоритмов организации доступа к ним. Данные или структуры данных в таких базах могут быть организованы в виде кольца, В-дерева, квадратичных структур и т. д. Сами базы данных могут содержать указатели, подструктуры и другую негеометрическую информацию. Организация баз данных и алгоритмы доступа к ним выходят за рамки изложения данной книги. Однако многие конкретные реализации систем с машинной графикой предназначены для работы с простыми изображениями, для которых можно предложить простые структуры данных с простым способом доступа. Наиболее элементарной структурой данных, безусловно, является линейный список. Может показаться неожиданным, но оказалось, что с помощью этой структуры представляются достаточно качественные изображения.

Точки — основные строительные блоки графической базы данных, именно поэтому особый интерес представляют основные операции преобразования точек. Существуют три типа операций, в которых точки используются как графические примитивы: во-первых, перемещение пера и курсора, след курсора (в дальнейшем просто курсор), невидимая точка; во-вторых, изображение линии между двумя точками; в третьих, вывод точки в заданном месте. В общем случае местоположение точки можно описать двумя способами: с помощью абсолютных и относительных (инкрементальных) координат. В системе относительных координат позиция точки определяется путем задания ее положения относительно предыдущей точки. Все программное обеспечение машинной графики базируется на этих основных понятиях и операциях. Более подробно основы построения программ машинной графики излагаются в разд. 1-22 и в приложении А.

Местоположения точки могут быть описаны с помощью вещественных (с плавающей точкой) или целых чисел. В случае применения целых чисел возникает множество проблем, связанных с ограниченной длиной слова компьютера. Для представления целых координат, как правило, используется полное машинное слово. Наибольшее целое число, занимающее полное машинное слово и необходимое для представления величин со знаком из положительного и отрицательного диапазонов, равно $2^{n-1} - 1$, где n — количество битов в слове. Для 16-битового слова это число равно 32 767. Для многих приложений этого бывает вполне достаточно, однако возникают серьезные проблемы, когда требуются большие числа, которые невозможно представить с помощью одного слова компьютера. Эту трудность можно преодолеть, например, воспользовавшись относительными координатами

в случае, когда их величина не превышает 60 000. При этом в абсолютной системе координат устанавливаются точки с координатами (30 000, 30 000), а затем с помощью относительных координат (30 000, 30 000) строится точка с координатами (60 000, 60 000). Однако этот способ обладает значительным недостатком: попытка сохранить относительные координаты, превышающие максимально допустимую величину, вызывает переполнение. В большинстве компьютеров результатом целочисленного переполнения является генерация чисел с противоположным знаком и неверным абсолютным значением.

От этого недостатка можно избавиться с помощью однородных координат, что вызывает некоторые усложнения, уменьшение скорости выполнения и понижение качества разрешения изображения. Однако все эти недостатки оправдываются преимуществами, которые дают возможность представления больших чисел на компьютерах с ограниченной длиной слова. Именно по этой причине, а также в связи с другими преимуществами использования однородных координат, им уделяется основное внимание в данной книге.

В системе однородных координат n -мерное пространство представляется $n + 1$ измерением, т. е. трехмерные координаты, однозначно задаваемые тройкой (x, y, z) , выражаются через четыре координаты (hx, hy, hz, h) , где h — произвольный множитель. Если все координатное пространство в 16-битовом компьютере представляется числами меньше 32 767, то h будет равно 1, а координаты будут представляться путем прямых преобразований. Однако если хотя бы одна из координат превышает число 32 767, например, $x = 60 000$, то преимущества использования однородных координат становятся очевидны. В данном случае мы установили $h = 1/2$ и координаты точки теперь можно представить в 16-битовом компьютере как $(30 000, y/2, z/2, 1/2)$. Однако в этом случае неизбежно снижение разрешающей способности, так как, например, значения $x = 60 000$ и $60 001$ теперь представляются одинаковой однородной координатой. Фактически разрешение ухудшается для всех координат даже в том случае, если только одна из них превышает максимально допустимое для целых чисел значение.

1-4 ПРЕДСТАВЛЕНИЕ ПРЕДВАРИТЕЛЬНО ПОДГОТОВЛЕННОГО ИЗОБРАЖЕНИЯ

Данные, используемые для подготовки рисунка, как правило, не совпадают с данными, необходимыми для его визуализации (дисплейным файлом). В этом файле содержится некоторая часть, вид или отдельная сцена целого изображения, представленного в общей базе данных. Выводимое изображение обычно формируется путем поворота, переноса, масштабирования и вычисления различных проекций. Эти преобразования обычно выполняются с помощью матричных операций (4×4) с данными, представленными в однородных координатах (гл. 2 и гл. 3). В случае, когда требуется последовательность преобразований, матрица каждого отдельного преобразования последовательно применяется к точкам рисунка. Однако если таких точек очень много, данный метод становится неэффективным. Гораздо удобнее другой способ, состоящий в использовании одной результирующей матрицы комбинированных преобразований, полученной путем умножения матриц каждого отдельного преобразования. Такая матричная операция называется кон-



Рис. 1-3 Двумерное отсечение.

катенацией. После этого вектор координат всех точек изображения умножается на матрицу комбинированного преобразования размером 4×4 для получения результирующих точек. Данный подход позволяет значительно сократить время проведения сложных матричных операций над множеством точек.

Удаление невидимых линий или поверхностей, закраска, определение прозрачности, текстуры или световые эффекты – все эти операции могут быть выполнены еще до момента окончательной визуализации изображения. Если выводимое изображение определено не на всей базе данных, в ней необходимо выделить некоторую часть, соответствующую этому изображению. Данный процесс называется отсечением. Отсечения бывают дву- и трехмерные. В ряде случаев окно и объем отсечения могут иметь отверстия или быть неправильной формы. Стандартное отсечение по границам дву- или трехмерных областей часто реализуется аппаратно. Подробное обсуждение особенностей реализации эффектов отсечения выходит за рамки данного изложения. Более детальное описание приведено в работе [1-1].

С процессом визуализации изображений обычно связано два ключевых понятия: окно и поле зрения. Процесс вывода с использованием окон предусматривает извлечение части графических данных из базы путем их отсечения по границам окна. Программная реализация операций отсечения и вывода окна, как правило, требует значительного времени, что исключает возможность ее применения в интерактивных графических системах. С другой стороны, в сложных графических системах для этого используется специализированная аппаратура или микропрограммные средства. Задача отсечения изображения включает выделение и отбрасывание отрезков или их частей, лежащих вне окна. Отсеченные отрезки или их части не передаются на графическое устройство вывода.

На плоскости окно задается левым, правым, верхним и нижним ребрами ограничивающего прямоугольника. Координаты вершин ребер определяются в системе координат пользователя или в мировой системе координат, т. е. именно в

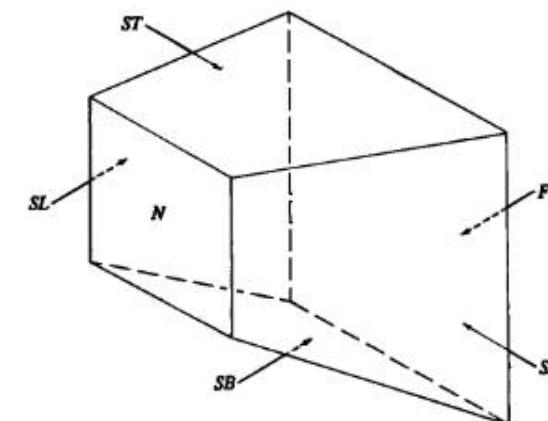


Рис. 1-4 Трехмерная пирамида видимости.

той системе, в которой хранится вся информация из графической базы данных. Как правило, координаты геометрических примитивов запоминаются в виде вещественных чисел.

Процесс отсечения значительно упрощается в случае так называемых регулярных окон, ребра которых параллельны осям координат. Нерегулярные отсекающие окна также представляют определенный интерес для многих конкретных приложений [1-1]. На рис. 1-3 показан процесс отсечения на плоскости. Линии удаляются, выводятся на экран полностью или частично в зависимости от их местоположения относительно границ окна. В трехмерном пространстве регулярное окно или отсекающий объем представляет собой прямоугольный параллелепипед (брюс) или, для перспективных видов, усеченную пирамиду (пирамиду видимости). На рис. 1-4 показана типичная пирамида видимости, ближайшая к наблюдателю грань помечена символом *N*, наиболее удаленная – символом *F*, а стороны образованы гранями *SL*, *SR*, *ST* и *SB*.

Поле зрения – это область на экране дисплея, предназначенная для размещения попавшего в окно изображения. Регулярное поле зрения на плоскости задается левым, правым, верхним и нижним ребрами ограничивающего прямоугольника. Поле зрения может быть определено в реальных физических координатах устройства, часто задаваемых в виде целых чисел. Координаты поля зрения могут быть нормализованы по некоторому произвольному диапазону, например, $0 \leq x \leq 1$, $0 \leq y \leq 1$ и задаваться в виде вещественных чисел. Содержание одного окна может быть выведено в нескольких полях зрения одного и того же устройства (рис. 1-5). Сохранение пропорций окна и поля (полей) зрения позволяет избежать искажений. Отображение на поле зрения графических данных, попавших после отсечения в окно, предусматривает выполнение операций переноса и масштабирования (приложение А).

Кроме графической информации, большинство изображений могут содержать алфавитно-цифровые и символьческие данные. Существуют два основных способа генерации символов: программный и аппаратный. Если при программной ге-

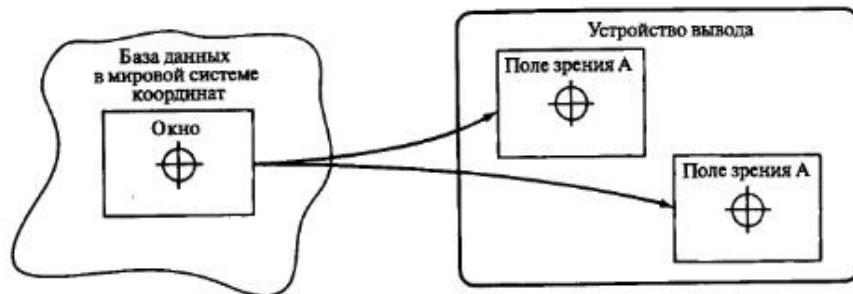


Рис. 1-5 Несколько полей зрения для одного окна.

нерации символов используются линии, то они отсекаются обычным образом. Это бывает необходимо в случае, когда символы подвергаются операции отсечения, а затем операции преобразования. Однако многие графические устройства имеют встроенные аппаратные генераторы символов, и в этом случае символы идентифицируются только своим кодом, а их изображение формируется непосредственно перед выводом на экран. Данная реализация намного эффективнее, но обладает меньшей гибкостью, чем программная генерация символов, так как не позволяет проводить отсечение и выполнять общие преобразования, ограничивающиеся в этом случае поворотом и масштабированием.

При использовании аппаратного генератора управляющая программа графического устройства должна установить размер, ориентацию и точку привязки символа или строки текста. После этого в дисплейный файл заносятся коды символов, имеющих установленные характеристики. Затем запускается генератор текстов, который интерпретирует строку текста, задавая аппаратуре всю информацию, необходимую для изображения каждого символа. После этого символы текста пересыпаются на устройство вывода.

1-5 ОРГАНИЗАЦИЯ ВЗАИМОДЕЙСТВИЯ С ИЗОБРАЖЕНИЕМ

После того как изображение на экране построено, возникает необходимость как-то взаимодействовать с ним или модифицировать его. Для этого был разработан ряд диалоговых устройств. Среди них можно назвать планшет, световое перо, рычаг, мышь, ручки для ввода скалярных величин, функциональные переключатели или кнопки и, разумеется, обычную алфавитно-цифровую клавиатуру. Прежде чем перейти к обсуждению этих физических устройств, рассмотрим функциональные возможности диалоговых графических устройств. Обычно насчитывается четыре или пять функциональных логических классов [1-2] – [1-4]. Логическими диалоговыми устройствами являются локатор, валютор, селектор и кнопка. Из-за широкой распространенности алфавитно-цифровой клавиатуры ее часто выделяют в пятый класс, называемый клавиатурой. В действительности клавиатуру можно концептуально и функционально считать набором кнопок.

Функцией локатора является выдача координатной информации в двух или трех измерениях. Обычно выдаваемые значения координат являются нормализо-

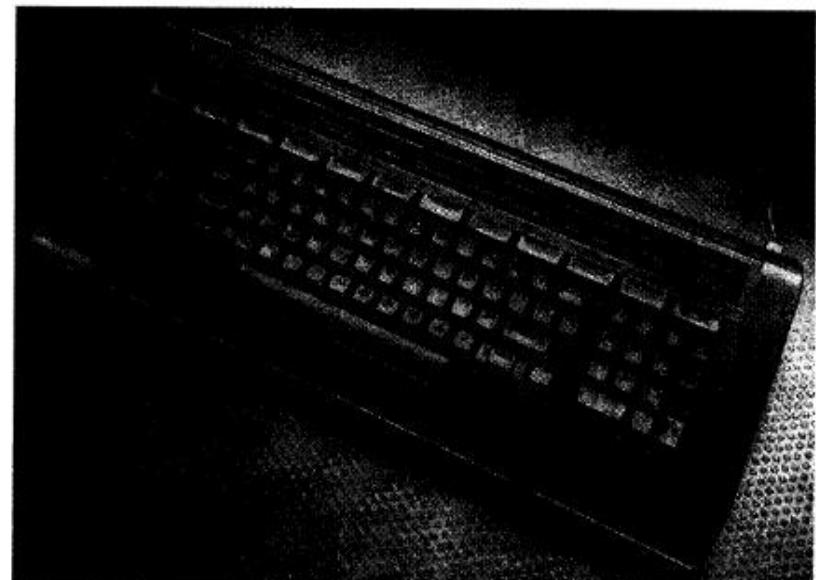


Рис. 1-6 Алфавитно-цифровая клавиатура. (С разрешения фирмы Evans & Sutherland Computer Corp.)

ванными и могут быть как относительными, так и абсолютными. Валютор применяется для ввода одиночной величины. Обычно это вещественное число между нулем и некоторым вещественным максимумом. Кнопка используется для выбора и активирования событий или процедур, управляющих сценарием диалога. Кнопка обычно представляет двоичную («включено» или «выключено») цифровую информацию. В функцию селектора входит идентификация или выбор объектов или подкартинок в выведенном изображении. Логическая клавиатура обрабатывает текстовую информацию. На рис. 1-6 показана типичная клавиатура.

Наиболее общим устройством класса локаторов является планшет (рис. 1-7). Планшеты можно использовать либо отдельно, либо в комбинации с графическим дисплеем на ЭЛТ. В первом случае их часто называют оцифровывателями. Сам по себе планшет состоит из плоской поверхности и карандаша (похожего на обычный карандаш), который используется для указания точки на поверхности планшета. Обычно положение карандаша распознается, если он находится на небольшом расстоянии от планшета. При работе с ЭЛТ обратная связь на экране обеспечивается с помощью небольшого символа (курсора), отслеживающего перемещение карандаша по поверхности планшета. В случае когда обратная связь используется в качестве отдельного оцифровывателя, она обеспечивается с помощью цифровых отсчетов.

Планшеты выдают координатную информацию в двух или трех измерениях. На рис. 1-8 показан трехмерный планшет. Выдаваемые планшетом значения представлены в координатах устройства. Они программно преобразуются в пространство координат пользователя. Обычное разрешение и точность составляют от 0.025 до 0.0025 см.

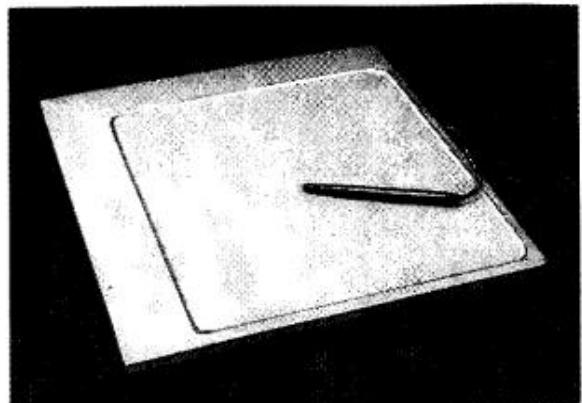


Рис. 1-7 Типичный планшет. (С разрешения фирмы Adage.)

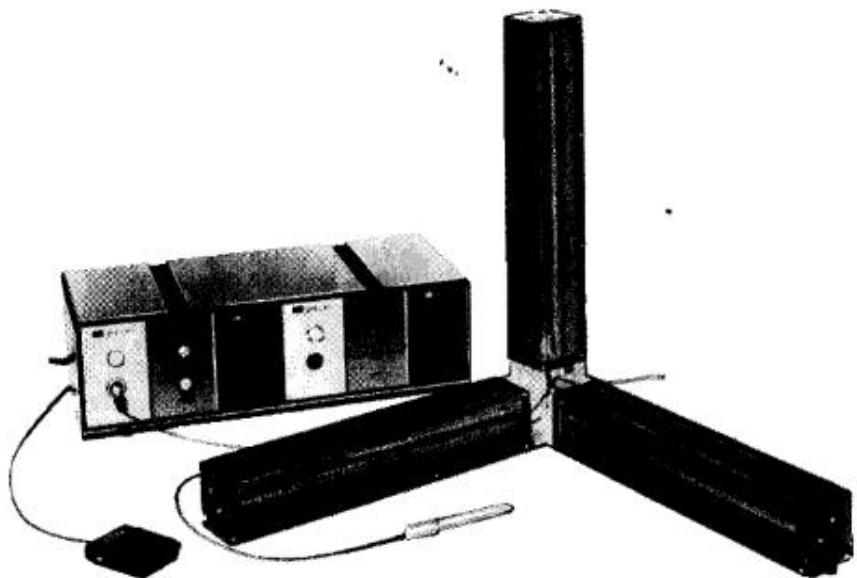


Рис. 1-8 Трехмерный звуковой планшет. (С разрешения фирмы Science Accessories Corp.)

При создании планшетов используется несколько различных принципов. В первом планшете RAND [1-5] применялась ортогональная сетка из отдельных проводов, расположенных под поверхностью. Каждый провод кодируется таким образом, чтобы карандаш, действующий как приемник информации, в каждой точке пересечения проводов получал уникальный цифровой код. Декодирование этого кода дает координаты карандаша x, y . Очевидными ограничениями на разрешение таких матрично кодируемых планшетов является плотность размеще-

ния проводов и способность приемника обрабатывать уникальный код. Точность ограничивается линейностью конкретных проводов, а также их параллельностью в двух ортогональных направлениях.

В одной достаточно интересной реализации планшета применяются звуковые волны. Карандаш используется для создания искры электрического разряда, вызывающей звуковую волну. Она движется во всех направлениях от карандаша по поверхности планшета, образуя круговой звуковой фронт. На его краях в перпендикулярных направлениях смонтированы два чувствительных ленточных микрофона. Координатные расстояния можно определить с помощью точного измерения времени, за которое волна проходит с карандаша до микрофонов. Данную методику можно распространить на три измерения (рис. 1-8).

Наиболее популярны конструкции планшета, в основе которых лежит электромагнитный принцип. При этом электрические импульсы проходят через сетку из проводов, расположенных под поверхностью планшета. Время, за которое чередующиеся импульсы, параллельные координатным осям x, y , проходят от края планшета до карандаша, определяется с помощью карандаша и соответствующих индикаторов. Эти данные легко преобразуются в координаты x, y .

Сенсорная панель похожа на планшет и относится к классу локаторов. В типичной сенсорной панели на двух смежных сторонах расположены источники света, а на двух противоположных смежных сторонах смонтированы светочувствительные элементы. Любой предмет, например палец, прерывая два ортогональных луча света, позволяет определить пару координат x, y . Из-за низкого разрешения этот прибор лучше всего использовать для грубых операций указания. С этой целью сенсорную панель часто монтируют поверх экрана ЭЛТ.

В таких локаторных устройствах, как рычаг, шар и мышь, в качестве делителя напряжения часто используются чувствительные переменные резисторы или потенциометры. Аналогичным образом реализуются и ручки для ввода скалярных величин, которые являются устройствами класса валюаторов. Точность всех этих устройств зависит от качества потенциометра и обычно колеблется от 0.1 до 10% всего диапазона измерения. Несмотря на то что разрешение потенциометра по существу бесконечно, его использование в цифровых системах требует аналого-цифрового (АЦ) преобразования. Обычно разрешение АЦ преобразователя изменяется в диапазоне от 8 до 14 бит, т. е. от 1/256 до 1/16384. Валюаторы также реализуются с помощью цифровых преобразователей углового положения, которые для каждого инкрементального угла поворота ручки выдают цифровой результат. Типичное разрешение — от 1/256 до 1/1024.

Типичным валюатором является рычаг (рис. 1-9). Подвижный рычаг обычно оборудуется двумя валюаторами — либо потенциометрами, либо преобразователями углового положения, смонтированными на основании прибора. Валюаторы выдают результаты, пропорциональные смещению ручки. В рычаг можно просто добавить третью степень свободы, используя, например, третий валюатор для измерения угла поворота. В качестве эха обратной связи рычага обычно употребляется графический курсор.

Шар во многом аналогичен рычагу (рис. 1-10). Чаще всего его можно встретить в радарных установках, например в системах управления воздушными перевозками. Сферический шар смонтирован в основании прибора, причем над по-



Рис. 1-9 Рычаг. (С разрешения фирмы Measurement Systems, Inc.)

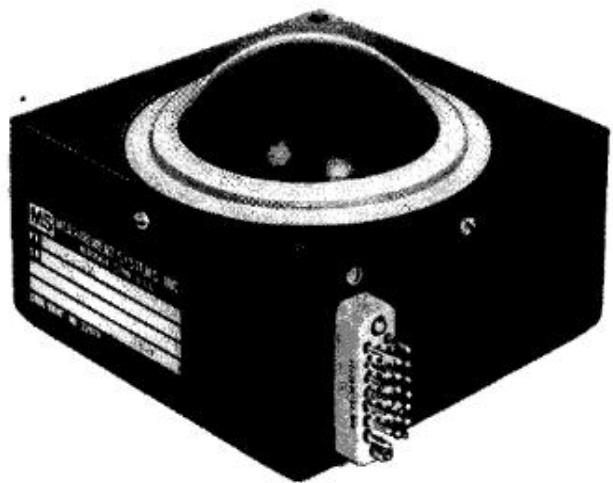


Рис. 1-10 Шар. (С разрешения фирмы Measurement Systems, Inc.)

верхностью возвышается только его часть. Шар может свободно вращаться в любом направлении. В основании смонтированы два валюатора (потенциометры или преобразователи углового перемещения), которые улавливают поворот шара и выдают результаты, пропорциональные относительным величинам углов. Кроме обычной обратной связи через курсор, пользователи получают тактильную обратную связь в виде скорости поворота или углового импульса вращения шара.

Рычаг имеет фиксированное положение с фиксированным началом координат. Мышь и шар, напротив, характеризуются только относительным положением ко-



Рис. 1-11 Мышь. (С разрешения фирмы Apple Computer, Inc.)

ординат. Обычная мышь состоит из шара, касающегося поверхности и расположенного в маленьком, легком футляре. При движении мыши по поверхности шар вращается и управляет осями двух валюаторов (потенциометров или преобразователей углового перемещения). Совокупное движение осей дает координаты x , y . Типичная мышь показана на рис. 1-11. Мыши можно поднять, переместить в любом направлении и поставить на поверхность, по-другому ориентировав. В этом случае изменяется система координат, в которой генерируются данные, т. е. меняется положение мыши, но не система координат, связанная с данными. Используемый для обратной связи курсор не движется, когда мышь не контактирует с поверхностью. Из-за проскальзывания шара возникает уменьшение точности. Совсем недавно в продаже появились устройства, работающие на оптическом или на магнитном принципе. В обоих случаях исключаются неточности, связанные с проскальзыванием.

Возможно, простейшими из валюаторов являются ручки для ввода скалярных величин. Такие ручки, показанные на рис. 1-12, являются по существу чувствительными вращающимися потенциометрами или точными цифровыми преобразователями углового положения, которые обычно объединяются в группы. Такие устройства особенно удобны при задании параметров для функций поворота, переноса или масштабирования.

Кнопки и функциональные переключатели, показанные на рис. 1-13, являются переключателями рычажного или нажимного действия. Они могут быть постоянно нажаты, постоянно отжаты либо иметь кратковременный контакт. Наиболее удобный тип функционального переключателя объединяет в себе обе эти возможности. Обычно можно программно управлять световыми индикаторами, указывающими, какие переключатели или кнопки в данный момент активны. Кнопки

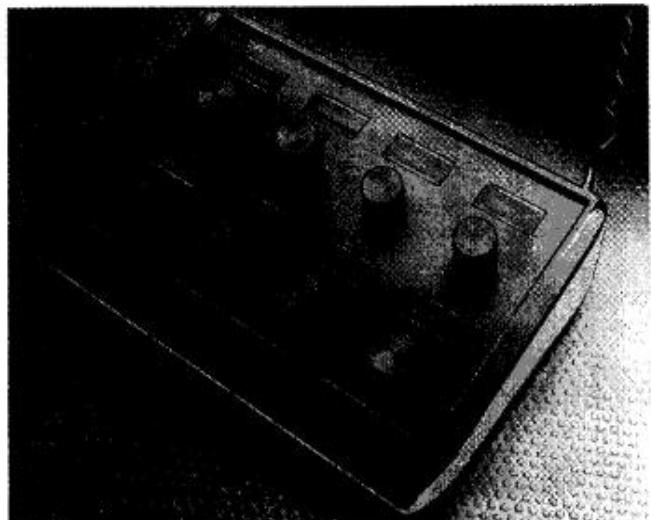


Рис. 1-12 Набор ручек для ввода скалярных значений. (С разрешения фирмы Evans & Sutherland Computer Corp.)

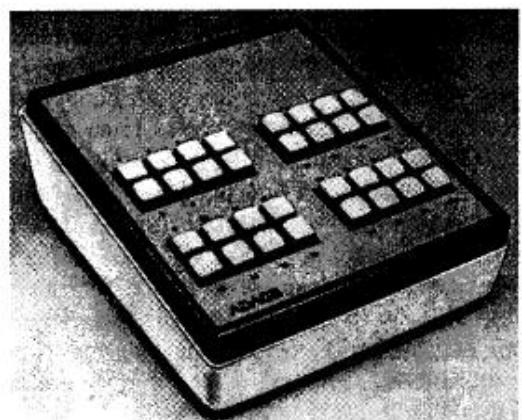


Рис. 1-13 Функциональные кнопки. (С разрешения фирмы Adage.)

и переключатели часто входят в состав других устройств. Например, на тонком конце карандаша планшета обычно располагается переключатель, срабатывающий при нажатии карандаша. На мыши также находится одна или несколько кнопок.

Единственным подлинным устройством указания является световое перо (рис. 1-14), которое содержит светочувствительный фотоэлемент и соответствующую электрическую цепь. Основная информация, передаваемая световым пером, связана с синхронизацией и зависит от того, что изображается на экране, и от порядка, в котором происходит его заполнение. Это исключает использование све-



Рис. 1-14 Схема светового пера.

тowego пера с дисплеями на запоминающей ЭЛТ (гл. 1-7). Его можно использовать только при дисплеях с регенерацией — векторных или растровых.

В тот момент, когда активизированное световое перо оказывается над той областью ЭЛТ векторного дисплея (гл. 1-8), в которой происходит рисование, дисплейному контроллеру посыпается сигнал. Он позволяет определить конкретную команду в дисплейном буфере, выполняемую в данный момент времени, что дает возможность идентифицировать отрезок, объект или подкартинку на экране. Используя графический курсор, можно применять световое перо в качестве локатора для векторного устройства с регенерацией изображения.

Так как в растровом дисплее изображение генерируется в фиксированной последовательности, световое перо используется для определения горизонтальной сканирующей строки (координаты y) и позиции в этой строке (координаты x). Отслеживая программу в контроллере, можно снова определить, какой отрезок, объект или подкартинка были указаны. Этот метод несколько усложняется в случае чересстрочной развертки (гл. 1-12). Приведенное выше описание означает также, что для растрового устройства световое перо может быть использовано скорее всего как локатор, а не как устройство указания.

Несмотря на то что для всех логических устройств существуют соответствующие физические устройства, в конкретной реализации некоторые из них могут отсутствовать. Таким образом, необходимо уметь моделировать логические интерактивные устройства ввода. На рис. 1-15 приведен пример, в котором световое перо используется для моделирования функций логической кнопки путем выбора из меню световых кнопок.

Планшет — одно из самых универсальных физических устройств. Его можно использовать как оцифровыватель для получения координат x , y . Вдобавок к этому оно прекрасно подходит для моделирования всяких логических интерактивных функций (рис. 1-16). Из этого же рисунка видно использование планшета в качестве локатора (позиция (а) на рис. 1-16). Используя графический курсор,



Рис. 1-15 С помощью выбора из меню световое перо используется для моделирования функции логической кнопки. (С разрешения фирмы Adage.)

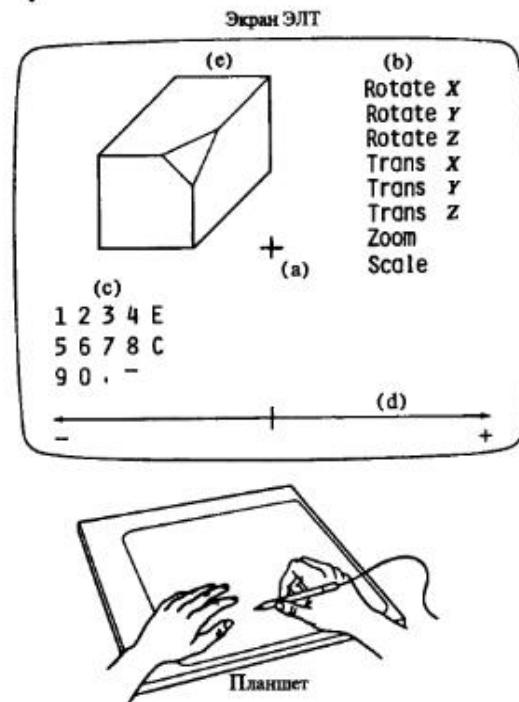


Рис. 1-16 Использование планшета для моделирования всех логических интерактивных функций: (а) локатора, (б) кнопки, (с) клавиатуры, (д) валюатора, (е) селектора.

можно реализовать функцию кнопки. Для этого с помощью планшета указываем курсором на кнопку меню (либо вблизи нее). Координаты, полученные с планшета, сравниваются с известными координатами x, y кнопок меню. Если обнаружено совпадение, то соответствующая кнопка активизируется (позиция (б) на рис. 1-16). Аналогичным образом на планшете можно реализовать клавиатуру (позиции (с) на рис. 1-16).

Отдельный валюатор обычно реализуется в сочетании с кнопкой. Конкретная функция для определения значения выбирается кнопкой, как правило, из меню. Затем валюатор можно смоделировать с помощью «числовой оси» (позиция (д) на рис. 1-16). Перемещение курсора вдоль оси генерирует координаты x, y , одна из которых интерпретируется как процентное отношение к диапазону валюатора.

Функцию указания можно реализовать с помощью локатора путем определения относительных координат x и y небольшого «окна указания». Данное окно становится затем графическим курсором, и для его позиционирования используется карандаш планшета. Координаты x, y каждого из интересующих отрезков, объектов или подкартинок сравниваются затем с текущим местоположением. Если они попадают в пределы окна, то данный элемент считается указанным. Для сложных изображений программная реализация данного алгоритма может оказаться медленной, но при аппаратной реализации заметной задержки не будет. Несмотря на то что световое перо нельзя использовать в качестве оцифровывателя, его, подобно планшету, тоже можно применять для моделирования всех логических интерактивных функций.

1-6 ТИПЫ ГРАФИЧЕСКИХ УСТРОЙСТВ

Существует много разнообразных устройств для вывода изображений, построенных с помощью машинной графики. Наиболее типичными являются электронно-лучевые трубы (ЭЛТ), плазменные дисплеи, перьевые граffопостроители, матричные, электростатические и лазерные печатающие устройства, устройства для фото и киносъемки. Кроме устройств визуализации большое развитие получают устройства сохранения изображения.

Запоминающие ЭЛТ с прямым копированием изображения (рисование отрезками), векторные дисплеи (рисование отрезками) и растровые сканирующие дисплеи с регенерацией (поточечное рисование) образуют три основных класса дисплеев на ЭЛТ. Развитие электронной технологии позволило использовать в одном дисплее несколько способов изображения. Класс устройств с плоским экраном образуют газоразрядные плазменные, электролюминесцентные дисплеи, дисплеи на жидкких кристаллах и дисплеи, построенные на основе светодиодов. При обсуждении различных дисплеев мы стоим на пользовательской или концептуальной точке зрения, т. е. в основном рассматриваем функциональные возможности дисплеев, а не особенности их электронной схемы.

1-7 ГРАФИЧЕСКИЕ ДИСПЛЕИ НА ЗАПОМИНАЮЩЕЙ ТРУБКЕ

Из всех дисплеев на ЭЛТ наиболее просто устроен дисплей на запоминающей ЭЛТ с прямым копированием изображения. Запоминающую ЭЛТ, называемую

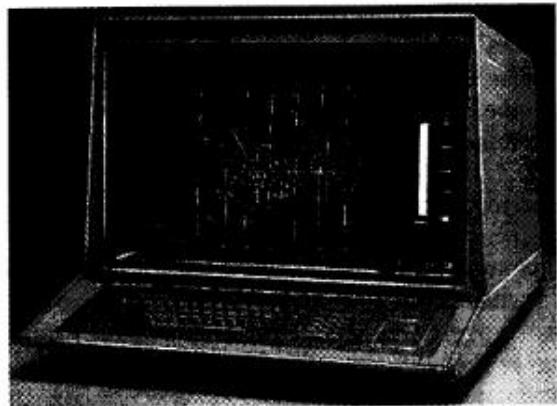


Рис. 1-17 Графический дисплей на запоминающей трубке. (С разрешения фирмы Tektronix.)

также бистабильной запоминающей трубкой, можно рассматривать как ЭЛТ, покрытую люминофором с длительным временем послесвечения. Линия или символ остаются на ней видимыми в течение длительного времени (до одного часа), прежде чем станут окончательно неразличимыми. На рис. 1-17 показан типичный дисплей такого типа. Чтобы нарисовать отрезок на дисплее, интенсивность электронного луча увеличивают до такой величины, которая вызывает запоминание следа луча на люминофоре. Для стирания изображения на всю трубку подают специальное напряжение, снимающее свечение люминофора. Экран вспыхивает и принимает исходное (темное) состояние. Весь процесс стирания занимает около 1/2 сек. Поскольку вспыхивает вся трубка, то стираются все отрезки и символы. Таким образом, стереть отдельные линии и символы невозможно, и нельзя осуществить динамическое изображение или анимацию. Иногда для обеспечения возможности ограниченной регенерации используется промежуточное состояние (режим рисования поверх изображения); см. ниже. В этом случае интенсивность электронного луча принимает значение меньше того порогового, которое вызывает запоминание, но достаточное для свечения люминофора. Поскольку в этом режиме изображение не сохраняется, для его видимости необходима постоянная перерисовка.

Дисплей на запоминающей трубке способен изображать фактически неограниченное количество векторов, а мерцание изображения в принципе невозможно. Его разрешение обычно составляет 1024×1024 адресуемых точек (10 бит) на экране 8×8 дюймов (ЭЛТ с диагональю 11 дюймов) или 4096×4096 (12 бит) либо на экране 14×14 дюймов (ЭЛТ с диагональю 19 дюймов), либо на экране 18×18 дюймов (ЭЛТ с диагональю 25 дюймов). По вертикали обычно видно только 78% адресуемой области.

Дисплей на запоминающей трубке — это векторный дисплей, или дисплей с произвольным сканированием. Иначе говоря, отрезок (вектор) может быть нарисован непосредственно из одной адресуемой точки в любую другую. Относительно легко, быстро и недорого можно получить твердую копию экрана. Дис-

плеи на запоминающей трубке в некоторой степени легче программировать, чем векторные или растровые дисплеи с регенерацией изображения. Дисплеи на запоминающих ЭЛТ можно объединять с микрокомпьютерами в сателлитные графические системы или графические терминалы. В последнем случае алфавитно-цифровая или графическая информация передается от главного компьютера в терминал через интерфейс. Обычно используется последовательный интерфейс, по которому за единицу времени передается только один бит информации, однако может применяться и параллельный интерфейс. Уровень интерактивности дисплея с запоминающей трубкой ниже, чем у векторного дисплея вследствие небольшой скорости интерфейса и неудовлетворительных характеристик стирания.

1-8 ГРАФИЧЕСКИЕ ДИСПЛЕИ С РЕГЕНЕРАЦИЕЙ ИЗОБРАЖЕНИЯ

В отличие от дисплея на запоминающей трубке в векторном (рисующем отрезки или векторы) дисплее с регенерацией изображения используется люминофор с очень коротким временем послесвечения. Такие дисплеи часто называют дисплеями с произвольным сканированием. Из-за малого времени послесвечения изображение на ЭЛТ за секунду должно многократно перерисовываться или регенерироваться. Минимальная частота регенерации должна составлять по крайней мере 30 кадров в секунду при рекомендуемой частоте от 40 до 50 кадр/с. Частота меньше, чем 30 кадр/с, приводит к мерцанию изображения, как это бывает при замедленном прокручивании кинофильма. На такое изображение неприятно смотреть, и его трудно использовать.

Для векторного дисплея с регенерацией кроме ЭЛТ требуются еще два элемента: дисплейный буфер и дисплейный контроллер. Дисплейный буфер — это участок непрерывной памяти, содержащей всю информацию, необходимую для вывода изображения на экран. Функция дисплейного контроллера состоит в циклической обработке полученной информации со скоростью регенерации. Сложность рисунка (количество изображаемых векторов) ограничивается двумя факторами — размером дисплейного буфера и скоростью дисплейного контроллера. Еще одним ограничением является скорость обработки геометрической информации, например скорость выполнения таких операций, как преобразование, отсечение и генерация текстов.

На рис. 1-18 представлены блок-схемы двух высокопроизводительных векторных дисплеев. В обоих случаях предполагается, что такие геометрические преобразования, как поворот, перенос, масштабирование, перспективное проецирование и отсечение, реализованы аппаратно в геометрическом процессоре. В первом случае (рис. 1-18а) геометрический процессор работает медленнее, чем это необходимо при регенерации применяемых на практике изображений (4000–5000 векторов). Таким образом, геометрические данные (экранные координаты векторов), посыпаемые центральным процессором (ЦП) графическому дисплею, обрабатываются до сохранения в дисплейном буфере. Следовательно, в нем содержатся только те инструкции, которые необходимы генератору векторов и символов. Обычно векторы задаются экранными координатами, а контроллер читает информацию из буфера и передает ее генератору векторов и символов. При до-

стижении конца дисплейного буфера контроллер возвращается на его начало, и цикл повторяется снова.

При использовании первой схемы реализуется концепция двойной буферизации и независимого изменения изображения и его регенерации. Так как в этой конфигурации геометрический процессор не успевает генерировать сложное новое или измененное изображение за один цикл регенерации, то дисплейный буфер делится на две части. В то время, когда измененное изображение обрабатывается и записывается в одну половину буфера, дисплейный контроллер регенерирует ЭЛТ из другой половины. При завершении изменения изображения буфера меняются ролями и процесс повторяется. Таким образом, новое или измененное изображение может генерироваться в каждый второй, третий, четвертый и т. д. циклы регенерации. Использование двойной буферизации предотвращает одновременный вывод части старого и части нового измененного изображения в течение одного и более циклов регенерации.

Во второй схеме (рис. 1-18б) геометрический процессор работает быстрее, чем необходимо для регенерации достаточно сложных изображений. В этом случае исходная геометрическая база данных, переданная из ЦП, сохраняется непосредственно в дисплейном буфере, а векторы обычно задаются в мировых координатах виде чисел с плавающей точкой. Дисплейный контроллер за один цикл регене-

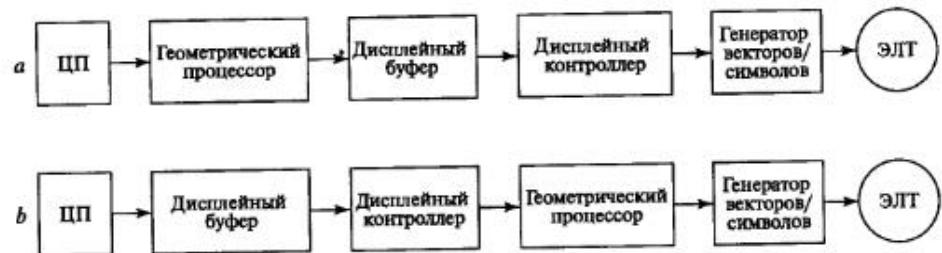


Рис. 1-18 Концептуальные блок-схемы векторных дисплеев с регенерацией изображения.

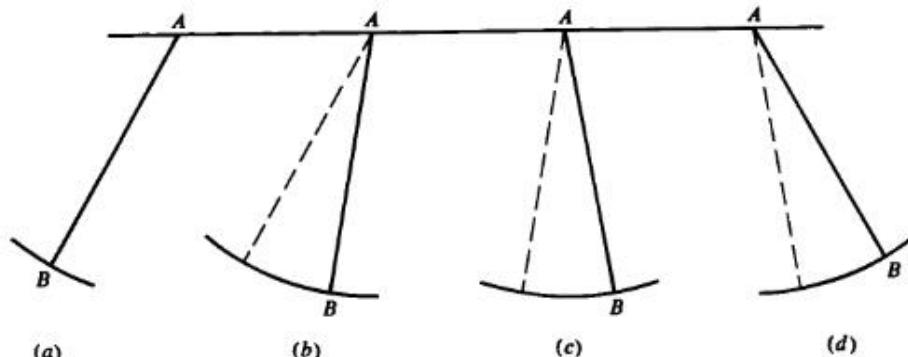


Рис. 1-19 Динамическое движение.

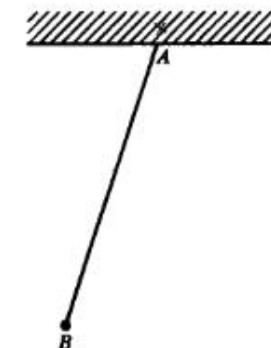


Рис. 1-20 Сегментация дисплейного буфера.

рации считывает информацию из буфера, пропускает ее через геометрический процессор и передает генератору векторов. При таком способе обработки геометрические преобразования выполняются «на лету» за один цикл регенерации.

При использовании любой конфигурации в дисплейном буфере содержатся команды рисования каждого вектора, символа и подкартинки. Следовательно, любой конкретный элемент изображения может быть изменен независимо от другого. Эта особенность вместе с малым временем послесвечения люминофора ЭЛТ позволяет выводить динамические изображения, что и показано на рис. 1.19, на котором представлена последовательность из четырех циклов. Отрезок, изображенный сплошной линией, — это отрезок, рисуемый для текущего цикла, штриховой линией изображен отрезок из предыдущего цикла. Между двумя циклами регенерации происходит изменение координат конца отрезка — точки B. Создается впечатление, что отрезок вращается вокруг точки A.

В большинстве случаев лишь часть изображения является динамически изменяемой. В действительности основная часть картинки остается статичной. Такое разделение наводит на мысль о необходимости сегментации дисплейного буфера, которая показана на рис. 1-20. Неподвижны, т. е. не изменяются от одного цикла регенерации к другому: горизонтальный отрезок, заштрихованная часть и буква A, используемые для показа опоры отрезка AB. В то же время для получения динамического изображения меняются положения конца отрезка AB и буквы B. Эти изолированные части базы данных помещены в отдельные сегменты дисплейного буфера. Для конфигурации, показанной на рис. 1-18а, геометрический процессор может пропускать статический сегмент в дисплейном буфере, так как он не изменяется, и это существенно сокращает работу процессора при изменении рисунка. В данном случае можно модифицировать изображение только в динамическом сегменте. Еще одним достоинством подобного метода является сокращение общего количества данных, передаваемых из ЦП в геометрический процессор при каждом изменении изображения.

Для конфигурации, представленной на рис. 1-18б, возможны разные типы сегментации. Напомним, что для данного случая база данных изображения сохраняется в дисплейном буфере в виде мировых координат, а обработка изображения

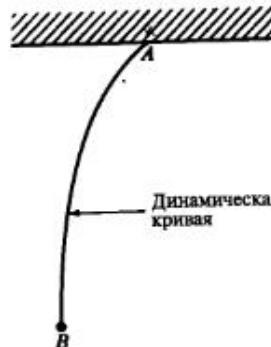


Рис. 1-21 Интеллектуальная сегментация дисплейного буфера.

происходит «на лету», за каждый цикл регенерации. Для картинки, показанной на рис. 1-20, в дисплейном буфере создаются два сегмента — статический и динамический. В любом случае обработка совершается мгновенно. Информацию в динамическом сегменте можно изменить с помощью функций, предоставляемых геометрическим процессором. Таким образом, модификация изображения может происходить локально в графическом устройстве, а для этого не требуется связь с ЦП. В конкретном случае на рис. 1-20 единственной функцией, необходимой для локального динамического изменения, является поворот вокруг точки *A*.

Для динамического изменения применительно к рис. 1-21 требуется связь с ЦП, т. е. необходимо некоторое интеллектуальное изменение изображения. Снова создаются два сегмента: статический, содержащий опорную линию, заштрихованную часть и букву *A*, и динамический, состоящий из кривой *AB* и буквы *B*. Предположим, что форма кривой *AB* должна изменяться от одного цикла регенерации к другому в зависимости от физических факторов. Это изменение выполняет прикладная программа в ЦП. Для модификации сегмента с динамически изменяющимся изображением должны быть переданы и сохранены в дисплейном буфере новые данные, например форма кривой.

Применение сегментации изображения не ограничивается только движением или мультилипликацией, хотя само понятие было введено с помощью примеров из этой области. Сегментировано может быть любое изображение, что особенно полезно для интерактивных графических программ. Понятие сегментации аналогично модульному программированию. Выбор модульных сегментов изображения, их размер и сложность зависят от конкретной прикладной области. Сложность элементов изображения изменяется от отдельных точек до полных описаний объектов. Типичный векторный дисплей с регенерацией изображения показан на рис. 1-22.

Чтобы показать, насколько важна скорость передачи или ширина полосы канала связи между ЦП и графическим устройством, рассмотрим затраты на интеллектуальное изменение кривой, описываемой 250 отрезками или точками, заданными тремя координатами. Пусть для представления чисел с плавающей точкой используются шесть значащих цифр (литер), и пусть каждая литера задается одним байтом (8 бит). Тогда при скорости регенерации 30 кадр/с и при усло-

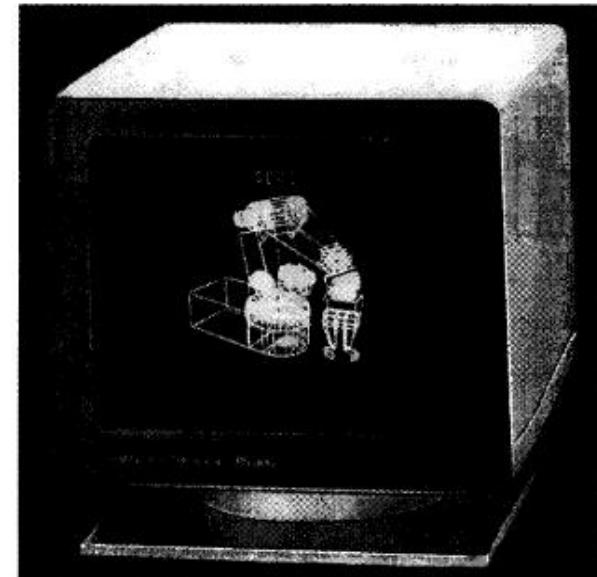


Рис. 1-22 Векторный дисплей с регенерацией изображения. (С разрешения фирмы Evans & Sutherland Computer Corp.)

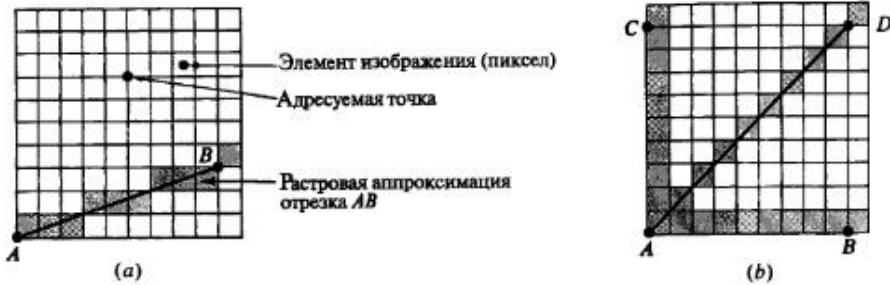
вии, что модификация производится в каждом цикле, требуемая ширина полосы связи составит

$$30[(\text{кол-во точек})(\text{кол-во коорд./точку})(\text{кол-во знач. цифр/точку})(\text{кол-во бит/литера})] \text{ или} \\ 30(250)(3)(6)(8) = 1\,080\,000 \text{ бит/с}$$

Таким образом, необходимая скорость передачи данных может легко превысить 1 Мбит/с. Для сложных трехмерных скульптурных поверхностей требуемая ширина полосы может возрасти в 10 раз, т. е. достигнуть 10 Мбит/с. В большинстве случаев для поддержки динамической графики в реальном масштабе времени такие скорости требуют применения между ЦП и графическим устройством параллельного интерфейса или интерфейса с прямым доступом в память.

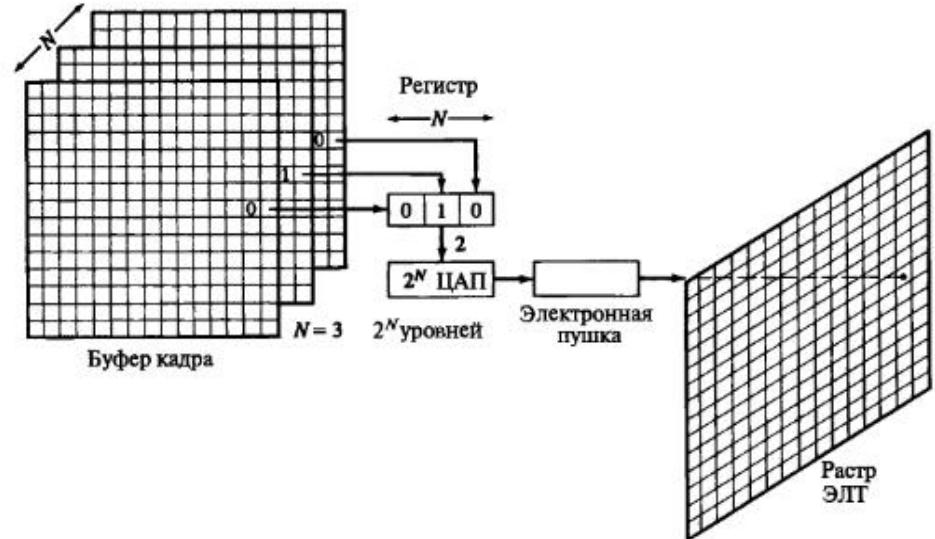
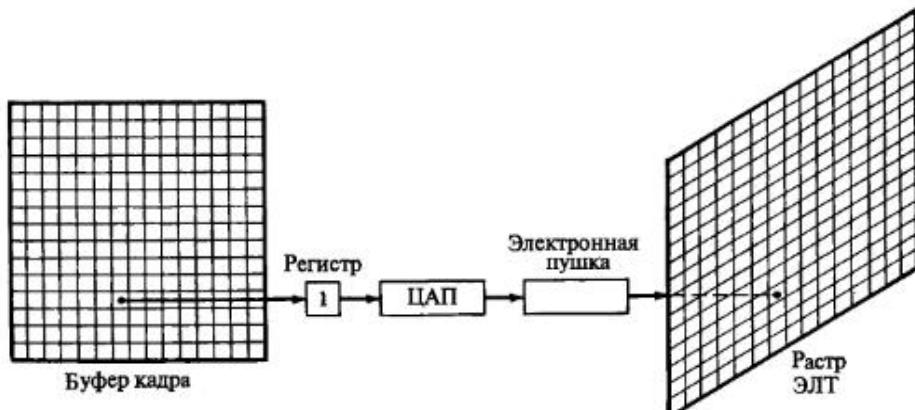
1-9 РАСТРОВЫЕ ГРАФИЧЕСКИЕ ДИСПЛЕИ С РЕГЕНЕРАЦИЕЙ ИЗОБРАЖЕНИЯ

Как дисплеи на запоминающих ЭЛТ, так и дисплеи с произвольным сканированием являются устройствами рисования отрезков, т. е. отрезок прямой может быть нарисован непосредственно из любой адресуемой точки в любую другую. Графическое устройство на растроевой ЭЛТ работает иначе. Раstroвое устройство можно представить в виде матрицы дискретных ячеек (точек), каждая из которых может быть подсвечена, т. е. оно является точечно-рисующим устройством. Невозможно, за исключением специальных случаев, непосредственно на-



рисовать отрезок прямой из одной адресуемой точки или пикселя в матрице в другую адресуемую точку или пиксель. Отрезок можно лишь аппроксимировать последовательностями точек (пикселов), близко лежащими к реальной траектории отрезка, что и проиллюстрировано на рис. 1-23 a . Как видно из рис. 1-23 b , отрезок прямой из точек получится только в случае горизонтальных, вертикальных или расположенных под углом 45° отрезков. Все другие отрезки будут выглядеть как последовательности ступенек. Это явление называется лестничным эффектом, или «зазубренностью». Методы, позволяющие устранить лестничный эффект обсуждаются в [1-1].

Чаще всего для графических устройств с растворной ЭЛТ используется буфер кадра. Буфер кадра представляет собой большой непрерывный участок памяти компьютера. Для каждой точки или пикселя в растре отводится как минимум один бит памяти. Эта память называется битовой плоскостью. Для квадратного раstra размером 512 × 512 требуется 2^{18} ($2^9 = 512$; $2^{18} = 512 \times 512$), или 262 144, бита памяти в одной битовой плоскости. Изображение в буфере кадра строится



побитно. Из-за того что бит памяти имеет только два состояния (двоичные 0 или 1), при одной битовой плоскости можно получить лишь черно-белое изображение. Битовая плоскость является цифровым устройством, тогда как растворная ЭЛТ – аналоговое устройство, для работы которого требуется электрическое напряжение. Поэтому при считывании информации из буфера кадра и ее выводе на графическое устройство с растворной ЭЛТ должно происходить преобразование из цифрового представления в аналоговый сигнал. Такое преобразование выполняет цифро-аналоговый преобразователь (ЦАП). Каждый пиксель буфера кадра нужно считать и преобразовать, прежде чем он будет отображен на растворной ЭЛТ. На рис. 1-24 приведена схема графического устройства с черно-белой растворной ЭЛТ, которое построено на принципе буферизации кадра в одной битовой плоскости.

Цвета или полутона серого цвета могут быть введены в буфер кадра путем использования дополнительных битовых плоскостей. На рис. 1-25 показаны схемы буфера кадра с N битовыми плоскостями для градации серого цвета. Интенсивность каждого пикселя на ЭЛТ управляется содержимым соответствующих пикселов в каждой из N битовых плоскостей. В соответствующую позицию регистра загружается бинарная величина (0 или 1) из каждой плоскости. Получившееся в результате двоичное число интерпретируется как уровень интенсивности между 0 и $2^N - 1$. С помощью ЦАП это число преобразуется в напряжение между 0 (темный экран) и $2^N - 1$ (максимальная интенсивность свечения). Таким образом можно получить 2^N уровней интенсивности. Рис. 1-25 иллюстрирует систему с тремя битовыми плоскостями для 8 (2^3) уровней интенсивности. Для каждой битовой плоскости требуется полный объем памяти при заданном разрешении раствора: например, буфер кадра с тремя битовыми плоскостями для раstra 512 × 512 занимает 7 864 32 (3 × 512 × 512) битов памяти.

Таблица 1-1 Комбинации цветов для простого 3-битового буфера кадра

	Красный	Зеленый	Синий
Черный	0	0	0
Красный	1	0	0
Зеленый	0	1	0
Синий	0	0	1
Желтый	1	1	0
Голубой	0	1	1
Пурпурный	1	0	1
Белый	1	1	1

Число доступных уровней интенсивности можно увеличить, незначительно расширив требуемую для этого память и воспользовавшись таблицей цветов, как это схематично показано на рис. 1-26. После считывания из буфера кадра битовых плоскостей получившееся число используется как индекс в таблице цветов, в которой должно содержаться 2^N элементов. Каждый ее элемент может иметь W бит, причем W может быть больше N . В этом случае можно получить 2^W значений интенсивности, но одновременно могут быть доступны лишь 2^N из них. Для получения дополнительных значений интенсивности необходимо изменить или перезагрузить таблицу цветов.

Поскольку существует три основных цвета, можно реализовать простой цветовой буфер кадра с тремя битовыми плоскостями, по одной на каждый из цветов. Отдельная битовая плоскость управляет индивидуальной электронной пушкой для каждого из трех основных цветов, используемых в устройстве. Комбинации их трех основных цветов дают восемь цветов на экране ЭЛТ, которые вместе с соответствующими им двоичными кодами приведены в табл. 1-1. Схема простого цветового растрового буфера кадра показана на рис. 1-27.

Для каждой из трех цветовых пушек могут использоваться дополнительные битовые плоскости. На рис. 1-28 схематично показан цветовой буфер кадра с 8 битовыми плоскостями на каждый цвет, т. е. буфер кадра с 24 битовыми плоскостями. Каждая группа плоскостей управляет 8-разрядным ЦАП и может генерировать 256 (2^8) оттенков, или интенсивностей, красного, зеленого или синего цвета. Их можно комбинировать для получения 16 777 216 [$(2^8)^3 = 2^{24}$] возможных цветов. Это «полноцветный» буфер кадра.

Полноцветный буфер кадра можно еще увеличить путем использования групп битовых плоскостей в качестве индексов в таблицах цветов, как это показано на рис. 1-29. При N битах на цвет и W -разрядных элементах таблиц цветов одновременно могут быть показаны $(2^3)^N$ цветовых оттенков из палитры с $(2^3)^W$ возможных цветов. Например, при буфере кадра с 24 битовыми плоскостями ($N = 8$) и тремя 10-разрядными таблицами цветов ($W = 10$) могут быть получены 16 777 216 (2^{24}) цветовых оттенков из палитры 1 073 741 824 (2^{30}) цветов, т. е. около 17 млн. оттенков из палитры немногим менее чем 1 млрд. цветов. Несмотря на то что на рис. 1-29 схематически показаны три различных таблицы цветов, при небольшом количестве физических битовых плоскостей (не более 12) более предпочтительной является ситуация, когда изменение цветов осуществляется непрерывно в таблице с $(2^3)^N$ элементами.

1-9 Растворные графические дисплеи с регенерацией изображения

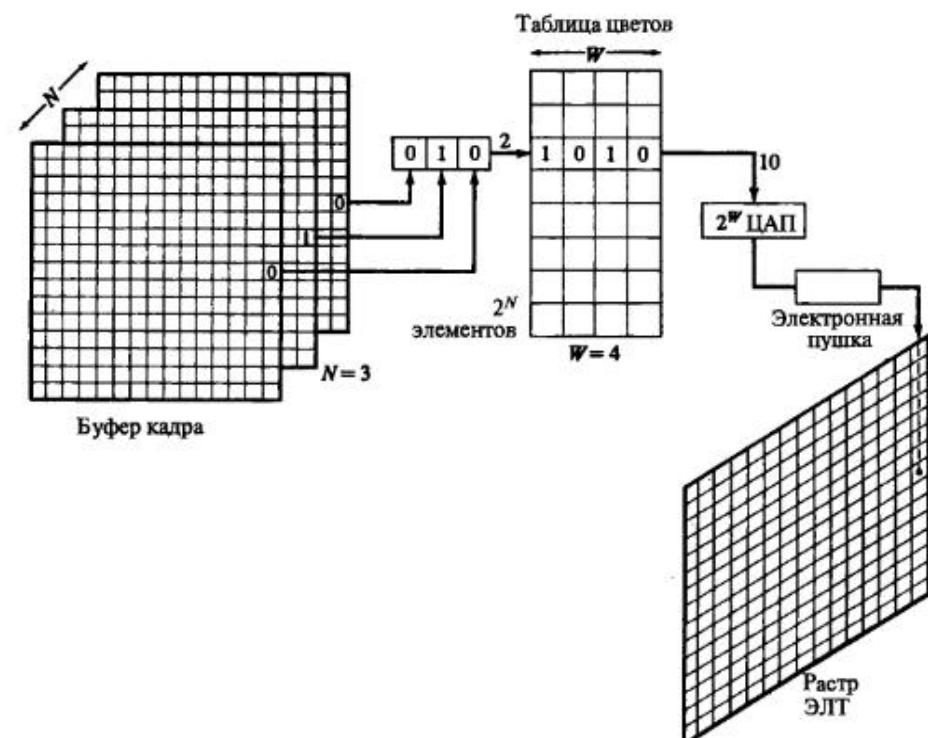
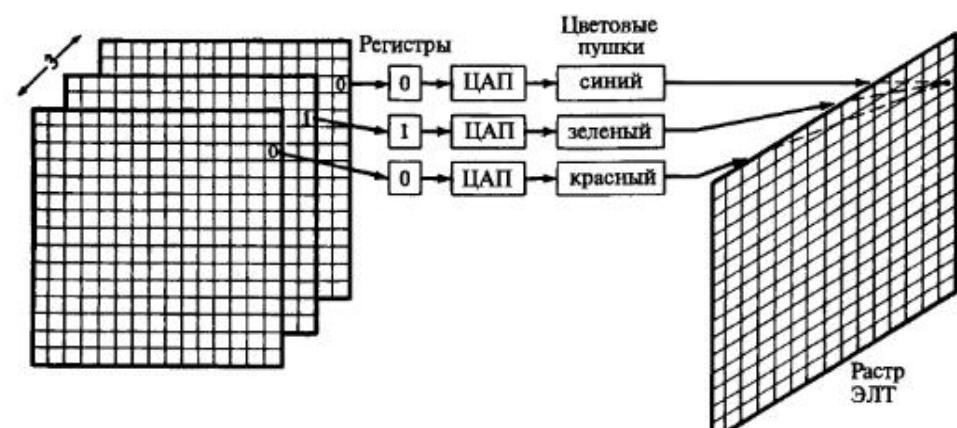
Рис. 1-26 Полутоновый черно-белый буфер кадра с N битовыми плоскостями и W -разрядной таблицей цветов.

Рис. 1-27 Простой цветовой буфер кадра.

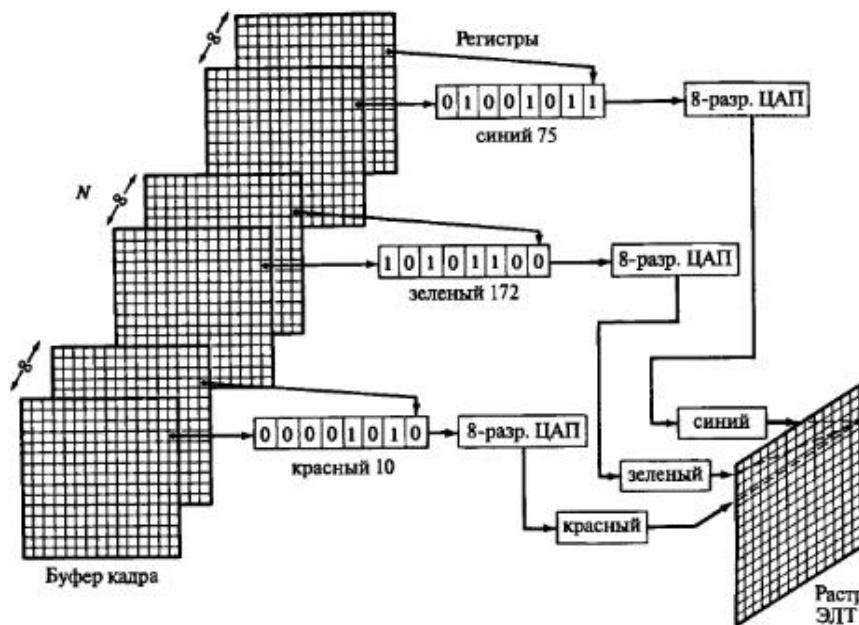


Рис. 1-28 Цветовой буфер кадра с 24 битовыми плоскостями.

Из-за большого количества пикселов в растровых графических устройствах трудно достигнуть производительности, необходимой для работы в реальном масштабе времени, а также для работы с приемлемой частотой регенерации, или смены кадра. Например, если среднее время доступа к каждому отдельному пикселиу равно 200 нс ($200 \times 10^{-9} \text{ с}$), то для организации доступа ко всем пикселям кадра размером 512×512 потребуется 0.0524 с , что эквивалентно частоте регенерации 19 кадров (картинок) в секунду, а это значительно меньше минимально необходимой частоты, равной 30 кадр/с. В буфере кадра размером 1024×1024 содержится немногим более 1 млн. бит (1 Мбит), и при среднем времени доступа 200 нс требуется 0.21 с для обращения ко всем пикселям, что составляет 5 кадр/с. Буфер кадра размером 4096×4096 содержит 16.78 млн. бит на каждую битовую плоскость! Доступ к ним займет 0.3 с. Для достижения скорости регенерации 30 кадров в секунду при таком растре требуется средняя эффективная скорость доступа 2 нс/пикセル. Напомним, что за этот небольшой период времени свет успевает пройти два фута.

Работа в реальном времени с растровыми графическими устройствами осуществляется путем одновременного доступа к группам по 16, 32, 64 и более пикселов. В случае цветового буфера кадра каждый пиксель может содержать до 32 бит, при этом все битовые плоскости для каждого пикселя доступны одновременно. При среднем времени доступа для каждой группы пикселов 1600 нс возможна работа в реальном времени для буферов кадров размером 512×512 и 1024×1024 .

Несмотря на то что для растровых устройств труднее достигнуть производительности, необходимой для работы в реальном масштабе времени с приемлемыми скоростями регенерации, чем на векторных дисплеях с регенерацией изображе-

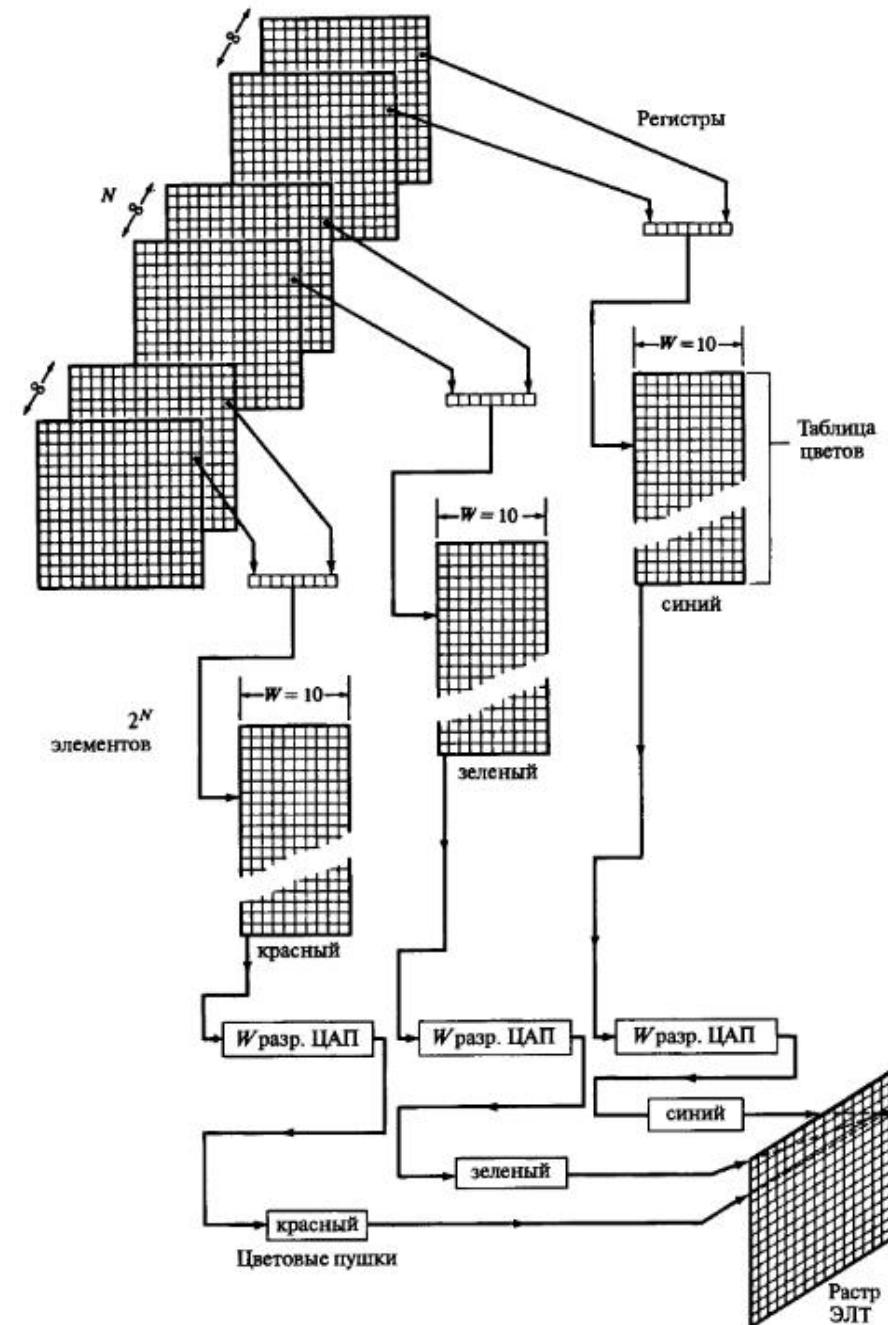


Рис. 1-29 Цветовой буфер кадра с 24 битовыми плоскостями и 10-разрядной таблицей цветов.

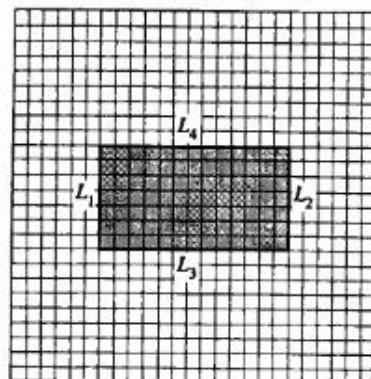


Рис. 1-30 Изображение сплошной фигуры на растровом графическом устройстве.

ния, на них легче изображать сплошные фигуры с плавными переходами цветов. Как показано на рис. 1-30, растровое представление сплошной «полигональной» фигуры концептуально достаточно просто. Здесь представление сплошной фигуры, ограниченной отрезками L1, L2, L3, L4, достигается установкой всех пикселов внутри ограничивающего многоугольника в соответствующий цвет в буфере кадра. Алгоритм «растровой развертки» сплошной области обсуждается в работе [1-1].

1-10 УСТРОЙСТВО ЭЛЕКТРОННО-ЛУЧЕВОЙ ТРУБКИ

Описанный выше буфер кадра сам по себе не является устройством вывода, он просто применяется для хранения рисунка. Наиболее часто в качестве устройства отображения, используемого с буфером кадра, выступает видеомонитор. Для того чтобы понять принципы работы растровых дисплеев и в некоторой степени векторных устройств с регенерацией, нужно иметь определенное представление об ЭЛТ и видеотехнике.

На рис. 1-31 схематично показана ЭЛТ, используемая в видеомониторах. Катод (отрицательно заряженный) нагревается до тех пор, пока возбужденные

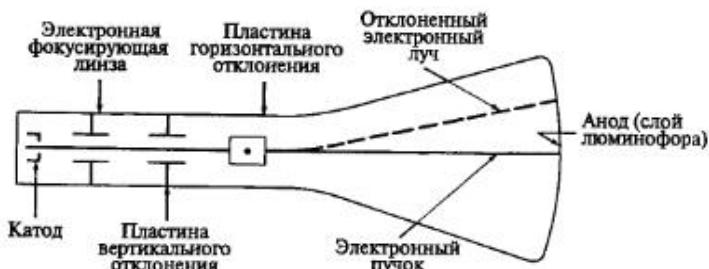


Рис. 1-31 Электронно-лучевая трубка.

электроны не создадут расширяющегося облака (электроны отталкиваются друг от друга, так как имеют одинаковый заряд). Эти электроны притягиваются к сильно заряженному положительному аноду. На внутреннюю сторону расширенного конца ЭЛТ нанесен люминофор. Если бы электронам ничто не препятствовало, то в результате их воздействия на люминофор весь экран ЭЛТ засветился бы ярким светом. Однако облако электронов с помощью электронных линз фокусируется в узкий, строго параллельный пучок, и теперь электронный луч дает одно яркое пятно в центре экрана. Луч отклоняется, или позиционируется, влево или вправо от центра и(или) выше или ниже центра с помощью усилителей горизонтального и вертикального отклонения.

Именно в данный момент проявляется отличие векторных дисплеев — как с запоминанием, так и с регенерацией — от растровых. В векторном дисплее электронный луч может быть отклонен непосредственно из любой произвольной позиции в любую другую позицию на экране ЭЛТ (аноде). Поскольку люминофор нанесен на экран сплошным слоем, то в результате получается почти идеальная прямая. В отличие от этого в растровом дисплее луч может отклоняться только в строго определенные позиции на экране, образующие своеобразную мозаику. Эта мозаика и составляет видеоизображение.

1-11 УСТРОЙСТВО ЦВЕТНОЙ РАСТРОВОЙ ЭЛТ

Цветная растровая ЭЛТ похожа на обычную черно-белую ЭЛТ, принцип действия которой был описан в предыдущем разделе. Данная трубка содержит три электронные пушки, по одной на каждый основной цвет: красный, зеленый и синий. Электронные пушки часто объединены в треугольный блок, соответствующий подобному треугольному блоку точек красного, зеленого и синего люминофоров на экране ЭЛТ (рис. 1-32). Для того чтобы электронные пушки возбуждали только соответствующие им точки люминофора (например, красная пушка возбуждала только точку красного люминофора), между пушками и поверхностью экрана помещена перфорированная металлическая решетка — так называемая теневая маска стандартной ЭЛТ. Отверстия в маске образуют такие же треуголь-

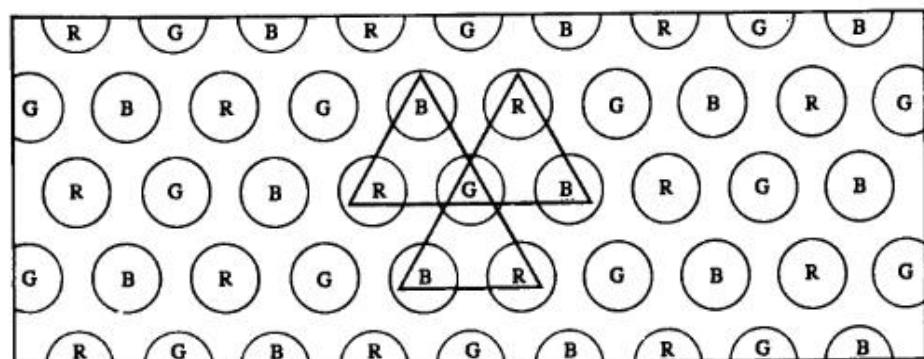


Рис. 1-32 Точечный люминофорный растр для ЭЛТ с теневой маской.

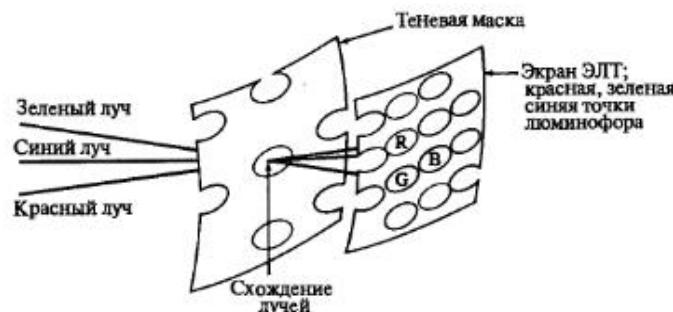


Рис. 1-33 Расположение электронной пушки и теневой маски цветной ЭЛТ.

ные блоки, как и точки люминофора. Расстояния между отверстиями называются шагом. Цветовые пушки расположены таким образом, что их лучи сходятся и пересекаются в плоскости теневой маски (рис. 1-33). После прохождения через отверстие красный луч, например, защищен или маскирован от пересечения с зеленой или синей точками люминофора. Он может пересечь лишь красную точку. Изменяя интенсивность электронного луча для каждого основного цвета, можно получить различные оттенки. Комбинация этих оттенков дает большое количество цветов для каждого пикселя. Обычно в дисплее с высоким разрешением на каждый пиксель приходится от двух до трех цветовых триад.

1-12 СИСТЕМЫ С ТЕЛЕВИЗИОННЫМ РАСТРОМ

Процесс преобразования хранящегося в буфере кадра растрового изображения в упорядоченный набор точек на телеэкране называется растровой разверткой. Сканируемый набор точек и частота воспроизведения основаны как на особенностях визуального восприятия, так и на принципах электроники. Системе визуального восприятия человека для рассмотрения элементов картинки требуется конечное время. Однако обеспечить впечатление непрерывности позволит только такой интервал, который настолько мал, что инерция зрительного восприятия перекроет мерцание. На мерцание влияет ряд факторов, включая яркость изображения и свойства конкретного люминофора, используемого для покрытия экрана ЭЛТ. Опыт показывает, что для практических целей минимальной скоростью вывода или изменения изображения является 25 кадр/с при условии, что минимальная скорость регенерации или воспроизведения в два раза больше, т. е. 50 кадр/с. Аналогичная ситуация имеет место и при демонстрации кинофильма. При этом оказывается 24 кадр/с, но каждый кадр показывается дважды, и в результате получается эффективная скорость воспроизведения 48 кадр/с. Таким образом, для фильма скорость изменения равна 24 кадр/с, а скорость регенерации — 48 кадр/с. В телевидении тот же эффект достигается с помощью метода чересстрочной развертки.

Телевидение использует метод растрового сканирования. В американской стандартной видеосистеме используются в совокупности 525 горизонтальных строк с кадровым, или видовым, отношением 4:3, т. е. высота видовой области равна трем

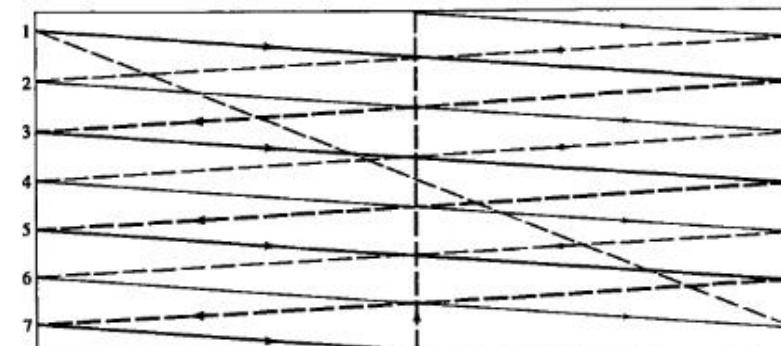


Рис. 1-34 Схема чересстрочной развертки для растра из семи строк. Нечетное поле начинается со строки 1. Штриховой линией обозначен горизонтальный обратный ход луча. Вертикальный обратный ход луча для нечетного поля начинается внизу в центре, а для четного поля — внизу справа.

четвертям ее ширины. Скорость воспроизведения, или смены кадра, составляет 30 кадр/с. Однако каждый кадр делится на два поля, каждое из которых содержит по половине картинки. Эти два поля чередуются или перемежаются. Они попеременно показываются через каждые 1/60 с. Одно поле содержит все строки с нечетными номерами (1, 3, 5, ...), а другое — с четными (2, 4, 6, ...). Сканирование начинается в верхнем левом углу экрана с нечетного поля. Каждая строка в поле сканируется или представляется слева направо. В то время как электронный луч движется поперек экрана слева направо, он также перемещается вертикально вниз, но с намного меньшей скоростью. Таким образом, «горизонтальная» сканирующая строка в действительности немного наклонена. При достижении правого края экрана луч становится невидимым и быстро возвращается к левому краю. Такой горизонтальный возврат луча обычно занимает около 17% времени, отведенного для одной сканирующей строки. Затем этот процесс повторяется со следующей нечетной строкой. Так как половина от 525 равна 262.5 строки, то при завершении сканирования поля нечетных строк луч окажется в центре нижней строки экрана (рис. 1-34 и рис. 1-35). Луч после этого быстро переводится в центр верхней стороны экрана. Так осуществляется вертикальный перевод луча для нечетного поля. Время, затрачиваемое на него, эквивалентно времени, затрачиваемому на сканирование 21 строки. Затем показывается поле четных строк, после чего луч оказывается в нижнем правом углу. Перевод луча для поля четных строк возвращает его в верхний левый угол, и весь процесс повторяется снова. Таким образом, показываются два поля для каждого кадра, т. е. 60 полей в секунду. Данный метод существенно уменьшает мерцание, так как глаз воспринимает частоту воспроизведения полей.

Хотя в принятой в США стандартной видеосистеме предусмотрено 525 строк, на самом деле видимы только 483 строки, так как время, затрачиваемое на сканирование 21 строки, уходит на вертикальный перевод луча¹. В течение этого

¹ Во многих графических устройствах с растровым сканированием это время используется для обработки другой информации.

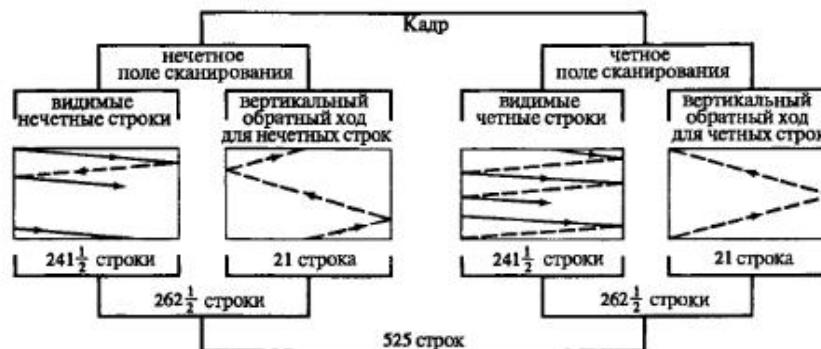


Рис. 1-35 Схема 525-строчного стандартного кадра.

времени электронный луч невидим или отключен. Время, отпущенное на каждую сканирующую строку, легко подсчитывается для частоты воспроизведения 30 кадр/с:

$$1/30 \text{ с/кадр} \times 1/525 \text{ кадр/строка} = 63,5 \text{ мкс/строка.}$$

Так как примерно 10,5 мкс расходуется на горизонтальный перевод луча, то вывод видимой части каждой строки должен быть завершен за 53 мкс. В строке при нормальном соотношении сторон видеообласти, равном 4:3, находятся 644 пикселя. Таким образом, время, отпущенное на считывание и вывод пикселя, равно

$$53 \text{ мкс/строка} \times 1/644 \text{ строка/пиксель} = 82 \text{ нс.}$$

Во многих растровых дисплеях, построенных на основе буфера кадра, применяется разрешение 512 пикселов на строку. Считывание и вывод пикселов при таком разрешении происходит примерно за 104 нс. Аналогичные результаты получаются для 625-строчной видеосистемы с частотой воспроизведения 25 кадр/с, используемой в Великобритании и большинстве стран Европы.

Метод чересстрочной развертки не является абсолютно необходимым при выводе изображения, однако изображение без чередования строк будет несовместимым со стандартным телевизионным приемником. При отсутствии чересстрочной развертки придется для устранения мерцания увеличить частоту воспроизведения до 60 кадр/с, а это, безусловно, в 2 раза сокращает время для обработки пикселя. Более высокое разрешение по числу строк и количеству пикселов в строке также уменьшает это время; например, при разрешении 1024 × 1024 на считывание и вывод пикселя отводится в 4 раза меньше времени, чем при разрешении 512 × 512, — примерно 25 нс! В этом случае потребуется очень быстрая память для буфера кадра и такой же высокопроизводительный ЦПУ.

1-13 ДИСПЛЕИ С ПЛОСКИМ ЭКРАНОМ

Несмотря на то что в настоящее время электронно-лучевые трубы (ЭЛТ) в качестве устройств графического вывода находятся вне конкуренции, им присущ ряд

недостатков: большие габариты, тяжелый вес, хрупкость и ограниченные размеры экрана (обычно 30 дюймов по диагонали). На фоне этих и других ограничений ЭЛТ очень выигрышно выглядят плазменные дисплеи, принадлежащие к типу растровых устройств с регенерацией изображения.

Плазменные дисплеи по способу формирования изображения делятся на два принципиально различных класса устройств: активные (светоэмиссионные) и пассивные (основанные на модуляции света). К первым устройствам относятся ЭЛТ, дисплеи на газоразрядных панелях, электролюминесцентные и вакуумные дисплеи. Жидкие кристаллы и светодиоды представляют класс пассивных устройств.

Среди устройств с активной генерацией света для решения задач интерактивной машинной графики наиболее подходят плазменные и электролюминесцентные дисплеи, которые способны генерировать большие изображения с высокой разрешающей способностью. За исключением некоторых специальных приложений, например в авиации, где дисплеи на светоэмиссионных диодах обладают некоторыми преимуществами, дисплеи на жидких кристаллах являются наиболее удачными устройствами из класса систем с пассивной регенерацией света.

Как показано на рис. 1-36, изображение на плоской ЭЛТ получается после прохождения луча параллельно экрану и отражения под прямым углом. Использование эффекта отражения позволяет значительно уменьшить протяженность колбы ЭЛТ и, соответственно, габариты дисплея. ЭЛТ в форме плоской панели обладают теми же преимуществами, что и обычные ЭЛТ, но при значительно меньших габаритах. Большая длина «шейки» ЭЛТ ограничивает области ее применения. Использование вакуумных флюоресцентных дисплеев также в значительной степени сдерживается их габаритами.

Дисплеи на газоразрядных панелях, электролюминесцентные дисплеи и дисплеи на жидких кристаллах имеют несколько общих характеристик. Все они состоят из матрицы отдельных пикселов растра, в каждом из которых помещено некоторое устройство или вещество, которые под воздействием напряжения или тока начинают испускать свет либо модулировать падающий свет. Требуемое для этого напряжение или ток подается каждому пикселу через отдельный электрический вентиль, например тонкопленочный транзистор, диод или резистор типа

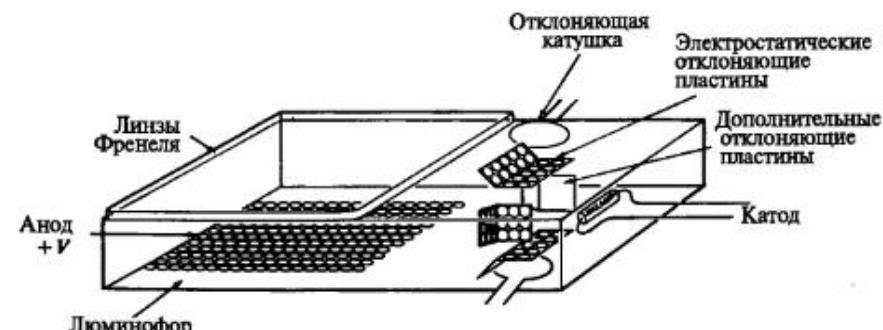


Рис. 1-36 Схема ЭЛТ.



Рис. 1-37 Типичный плазменный дисплей. (С разрешения фирмы Thomson CSF Electron Tubes & Device.)

металл—изолятор—металл, размещенный под каждым пикселом. Дисплеи, основанные на данном принципе, называются матричными.

Другой способ, получивший широкое распространение, основан на использовании адресации пикселов по строкам и столбцам, что позволяет значительно уменьшить количество вентилей или драйверов генерации изображения. Данный способ предполагает наличие у материала, из которого изготовлен дисплей, некоторого порога переключения. Эмиссия или модуляция света в таком элементе возможна только при превышении такого порога переключения. На соответствующую строку матрицы пикселов подается напряжение или ток, недостаточные для активации отдельного пикселя, а на соответствующий столбец подается другое напряжение или ток. Каждое в отдельности воздействие по строке или по столбцу ниже порогового значения, а в сумме они превышают его. По этой причине пиксели, не лежащие на пересечении нужной строки и столбца, оказываются в невозбужденном состоянии и только один пиксель получает напряжение или ток выше порогового значения становится активным.

При использовании такой адресации весьма желательно наличие для пикселов бистабильной памяти, в которой однажды активизированный пиксель сохраняет свое состояние, пока не произойдет его явное переключение. Бистабильная память исключает необходимость в постоянной регенерации пикселов, и, как следствие этого, не требуется внешней памяти для регенерации. Кроме того, значительно упрощается устройство дисплейного контроллера.

Принцип работы плазменного или газоразрядного дисплея предельно прост. Дисплей состоит из матрицы ячеек (растра), окруженных стеклянной оболочкой. Каждая ячейка заполнена газом (обычно неоном или смесью аргона с неоном) под давлением (ниже атмосферного). При действии достаточно высокого напряжения происходит диссоциация газа, т. е. атомы лишаются своих электронов. Такой газ в состоянии диссоциации называется плазмой, отсюда и возникло название — плаз-

менные дисплеи. При рекомбинации электронов происходит освобождение энергии в форме фотонов и газ начинает светиться с характерным оранжево-красным оттенком. На рис. 1-37 показан типичный плазменный дисплей.

Активация в плазменных дисплеях может быть типа AC (переменный ток) или DC (постоянный ток) либо представлять комбинацию AC/DC. На рис. 1-38 *a*, *b* и *c* приведены схемы плазменных дисплеев с активацией типа AC, DC и гибридной типа AC/DC соответственно. Дисплеи с типом DC являются более простыми, чем AC, и состоят из полой диэлектрической панели, содержащей полости с газом, которые помещены между панелями с матрицей проводников. Электронное поле прилагается непосредственно к газу. Дисплеи с типом DC обладают достаточно продолжительным циклом регенерации.

В дисплеях с типом AC слой диэлектрика размещается между проводником и газом. Таким образом, единственный путь соединения между газом и проводником — это емкость (конденсатор), и поэтому для получения диссоциированного газа требуется использование напряжения. Плазменные дисплеи типа AC имеют бистабильную память, что исключает необходимость в длительной регенерации. Эффект бистабильной памяти достигается при использовании низкого AC постоянного напряжения. Имеющейся емкости оказывается достаточно для поддержания напряжения активации токопроводящих пикселов, но недостаточно для активации непроводящих.

В гибридных дисплеях с типом активации AC/DC (рис. 1-38*c*) используется DC-напряжение для «первичного» газа, а для облегчения активации применяется AC-напряжение. Принципиальное преимущество гибридных AC/DC-плазменных дисплеев — это снижение сложности управляющей схемы. На практике находят применение как дисплеи больших размеров (3 × 3 метра с разрешением 4096 × 4096), так и системы высокого разрешения (100 пикселов на дюйм). На основе использования данных принципов можно также создавать цветные и полуточновые монохромные дисплеи.

В электролюминесцентном дисплее фосфоресцирующий материал излучает свет под действием электрического поля от AC или DC. В качестве фосфоресцирующего материала обычно используется сульфид цинка с добавками марганца, поэтому эти дисплеи излучают желтый свет. Для адресации пикселов используется такой же способ, как и в плазменных дисплеях. При подаче напряжения, превышающего пороговое значение, возбуждаются электроны марганцевой составляющей люминофора. Когда возбужденный атом возвращается на более низкую энергетическую орбиту, он испускает фотон, вызывающий характерный желтый цвет. Можно также получить хороший спектр серого цвета, так как уровень люминесценции управляется напряжением и частотой приложенного электрического поля. При использовании других типов добавок получаются различные цвета свечения покрытия, следовательно, из совокупности нескольких фосфоресцирующих слоев может быть сформирован цветной дисплей.

Люминофор представляет собой либо макроскопический субстрат, напыляемый в виде толстого слоя, либо тонкую пленку частиц толщиной в одну молекулу. Дисплеи с типом возбуждения тонкопленочного покрытия AC и DC получили наибольшее распространение в системах машинной графики. Принципиальная схема электролюминесцентного дисплея приведена на рис. 1-39, а на рис. 1-40

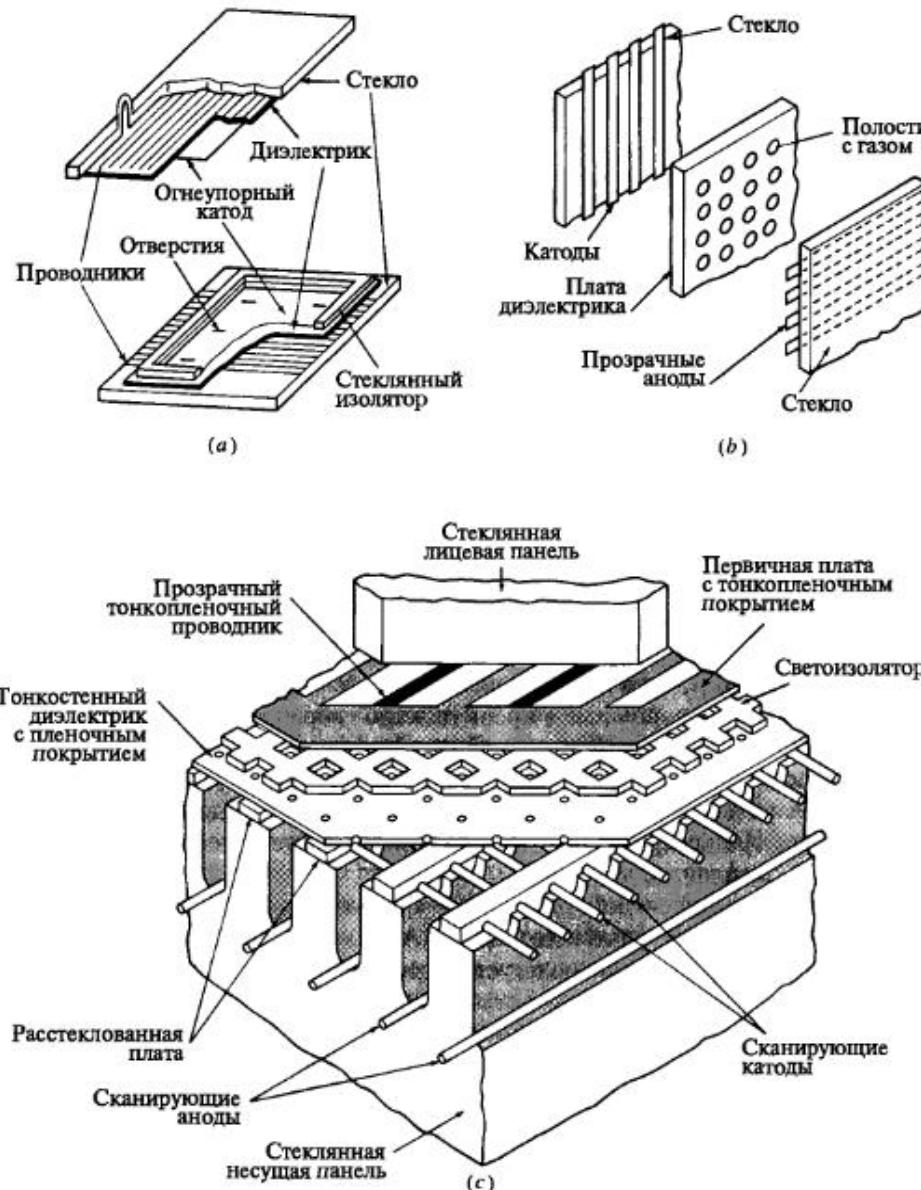


Рис. 1-38 Принципиальная схема плазменного газоразрядного дисплея: а) АС тип; б) DC тип; в) AC/DC тип.

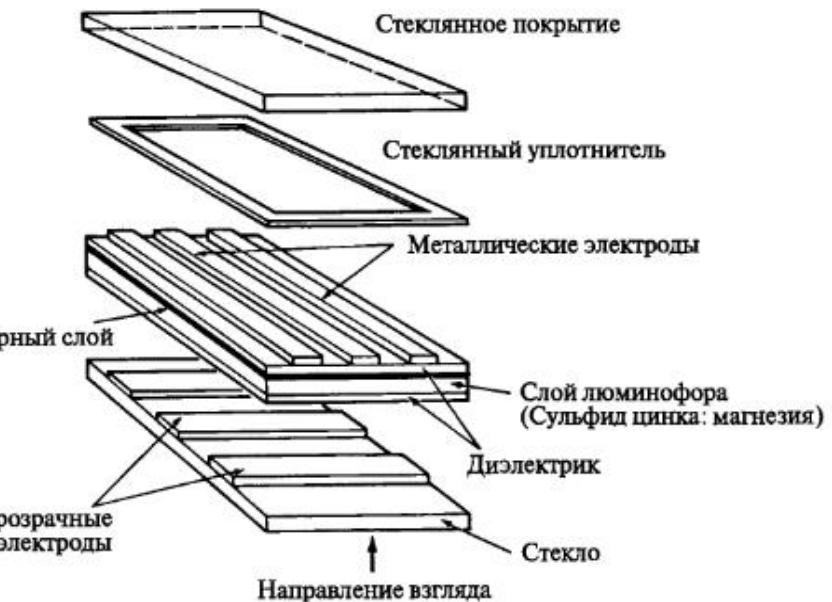


Рис. 1-39 Принципиальная схема электролюминесцентного дисплея типа АС.

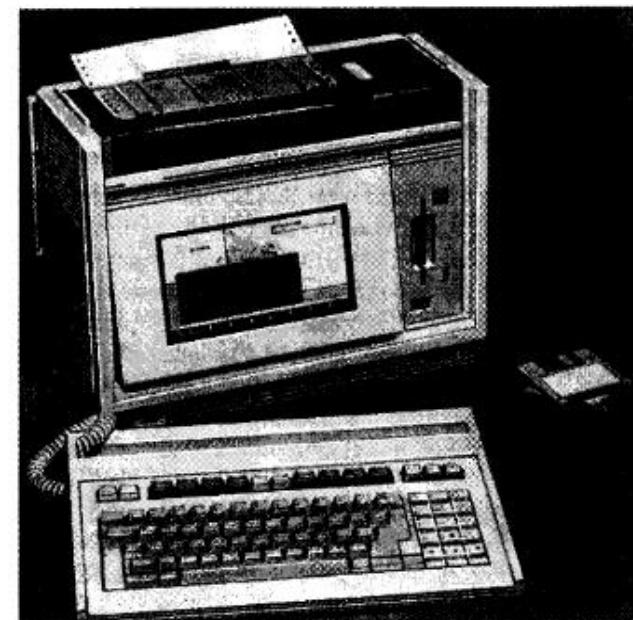


Рис. 1-40 Электролюминесцентный дисплей. (С разрешения фирмы Hewlett Packard.)

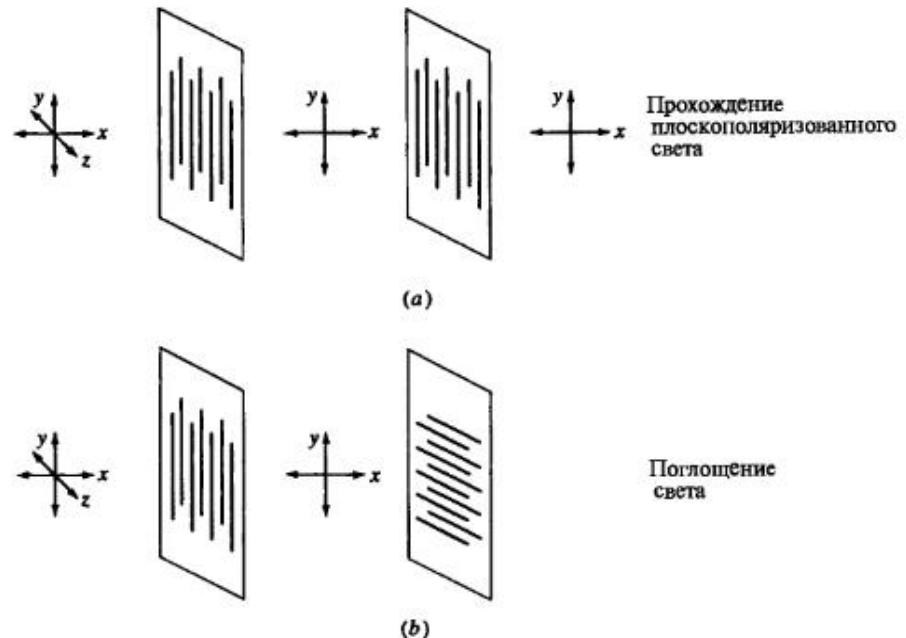


Рис. 1-41 Поляризация света: а) прохождение плоскополяризованного света; б) поглощение света.

изображен типичный дисплей такого типа. Из рисунка видно, что дисплей с разрешением 100 пикселов на дюйм имеет вполне компактные габариты.

Если плазменные и электролюминесцентные дисплеи представляют собой примеры устройств, использующих активную панель, то дисплеи на жидкых кристаллах являются образцом пассивной технологии, основанной либо на пропускании, либо на отражении падающего света. Для изменения характеристики падающего света используются свойства поляризации, которыми обладают некоторые органические вещества.

Рис. 1-41 наглядно иллюстрирует основные принципы поляризации света. Некогерентный свет (рис. 1-41а) проходит через первый (на рисунке левый) поляризатор и поляризуется по плоскости xy . Оси поляризации второго поляризатора также ориентированы в плоскости xy , поэтому свет беспрепятственно проходит дальше. На рис. 1-41б плоскость второго поляризатора повернута под прямым углом относительно первого, и поэтому свет, поляризованный в плоскости первого поляризатора, поглощается при прохождении через второй.

Органические вещества, способные сохранять стабильность в мезофазе при температуре между жидким и твердым состоянием, называются жидкими кристаллами. Различают три типа мезофазы: смектическую, нематическую и холестерическую.

В нематической фазе оси молекул жидкого кристалла располагаются параллельно, причем на их направление выравнивания влияют температура, поверхностное натяжение, давление и, что особенно важно для дисплеев, электрические

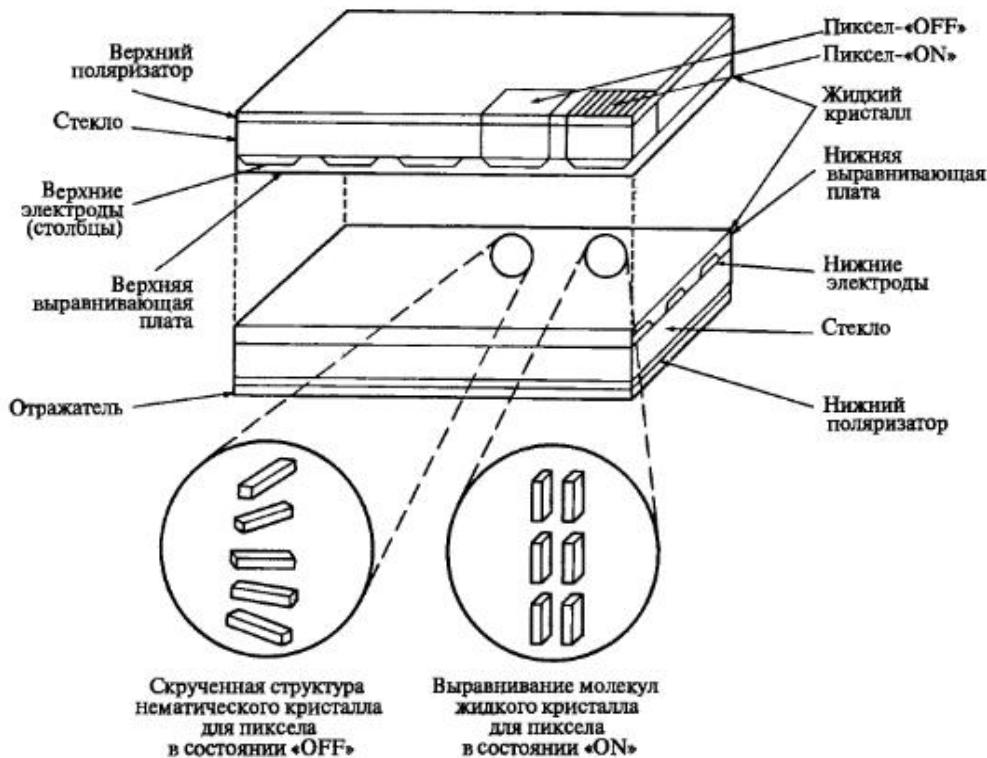


Рис. 1-42 Принципиальная схема дисплея на скрученных нематических жидкких кристаллах.

и магнитные поля. Оптические характеристики жидкого кристалла также достаточно чувствительны к этим воздействиям.

Ключевым моментом одного из возможных способов построения дисплеев на жидких кристаллах является создание «бутерброда» из нематических кристаллов со взаимоперпендикулярными осями. На рис. 1-42 приведено изображение принципиальной схемы отражающего нематического жидкокристаллического дисплея. Нижняя и верхняя плата бутерброда из жидких кристаллов имеют канавки, причем канавки на верхней плате направлены под прямым углом к канавкам нижней платы. Жидкие кристаллы прилегают к поверхности платы с соблюдением направлений канавок (рис. 1-42).

Дисплей имеет два плоских поляризатора, каждый из которых расположен на обеих сторонах бутерброда и повернут под прямым углом к другому. При затемненном пикселе или в состоянии поворота падающий на поверхность свет поляризуется первым поляризатором, затем проходит через жидкокристаллический бутерброд, где он поворачивается на прямой угол, далее идет сквозь второй поляризатор и затем отражается обратно, а пиксель становится подсвеченным.

Включение пикселя осуществляется путем подачи электрического поля на жидккий кристалл, что вызывает его переориентацию. Свет, падающий на па-

нель дисплея, поляризуется теперь первым поляризатором, проходит через жидкий кристалл без каких-либо изменений и оказывается полностью поглощенным вторым поляризатором. Пиксел становится темным. Жидкокристаллические дисплеи на повернутых нематических кристалах требуют для своей работы постоянной регенерации.

Бистабильные дисплеи на жидкких кристаллах могут быть построены на базе смектических кристаллов, две устойчивые ориентации молекул которых обладают различными оптическими свойствами, например уровнем отражения. Это позволяет достигать оптических контрастов за счет приложения импульсного напряжения, вызывающего изменение состояния молекул. Для дисплеев на кристаллах данного типа не требуется регенерации.

Использование цветных фильтров или фосфоров позволяет строить цветные дисплеи на основе повернутой нематической технологии или с использованием окрашивающих добавок. В дисплеях на кристаллах с добавками в основную жидкокристаллическую подложку вносятся двухцветные посторонние молекулы, спектральные характеристики которых используются для формирования различных цветов. Под действием электрического поля происходит изменение ориентации основных и привнесенных молекул, что дает возможность свету беспрепятственно проходить сквозь кристалл. На рис. 1-43 схематически показан типичный жидкий кристалл типа гость–хозяин. Отметим, что в этом случае требуется только один поляризатор. На основе данного кристалла можно построить также дисплеи, отражающие свет.

Дисплеи на жидкких кристаллах имеют адресацию типа строка–столбец либо содержат активные матричные элементы, например транзисторы или диоды в каждом пикселе. Обычная разрешающая способность данных дисплеев составляет 100 пикселов на дюйм.

1-14 ЭЛЕКТРОСТАТИЧЕСКИЕ ГРАФОПОСТРОИТЕЛИ

Электростатические графопостроители принадлежат к типу устройств с растровым сканированием (см. разд. 1-9, 1-12). Основной принцип их функционирования — это нанесение частиц краски на участки бумаги, обладающие определенным электрическим зарядом. На рис. 1-44 представлена обобщенная схема устройств данного типа. Специальное покрытие носителя изображения, удерживающее электростатический заряд, помещается около записывающей головки, которая содержит от 1 до 3 рядов пишущих узлов или перьев. Обычно на одном дюйме располагается от 72 до 400 таких перьев. Чем больше рядов пишущих узлов, тем выше плотность печати. При прохождении носителя под рядами перьев каждое из них наносит отрицательно заряженную точку. Далее носитель проходит через тоновый аппликатор, где положительно заряженные частицы краски притягиваются к точкам изображения, имеющим отрицательный заряд, в результате на носителе проступает видимое изображение. Затем частицы графитовой краски просушиваются и фиксируются на носителе. В качестве материала носителя изображения для графопостроителя может быть использована непрозрачная или полупрозрачная пергаментная бумага и прозрачная или матовая полиэфирная пленка.

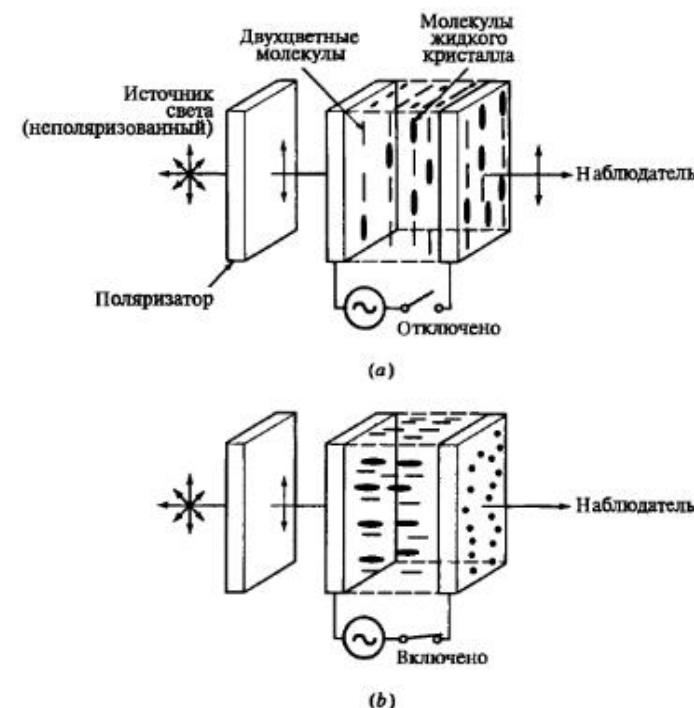


Рис. 1-43 Жидкие кристаллы типа гость–хозяин (закрашены). Молекулы ориентированы для блокировки (a) и для свободного прохождения (b).

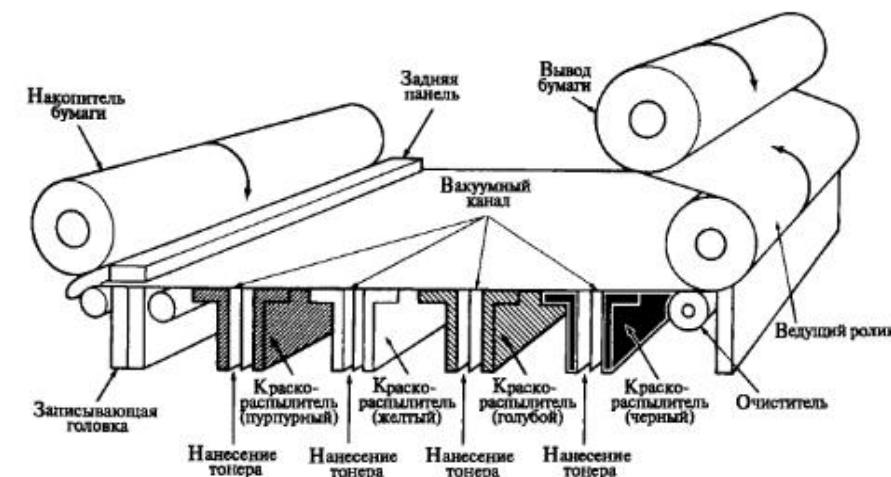


Рис. 1-44 Принципиальная схема цветного электростатического графопостроителя.

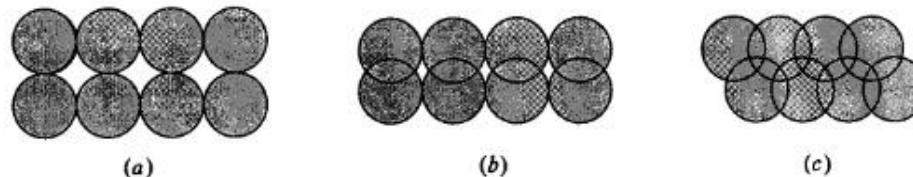
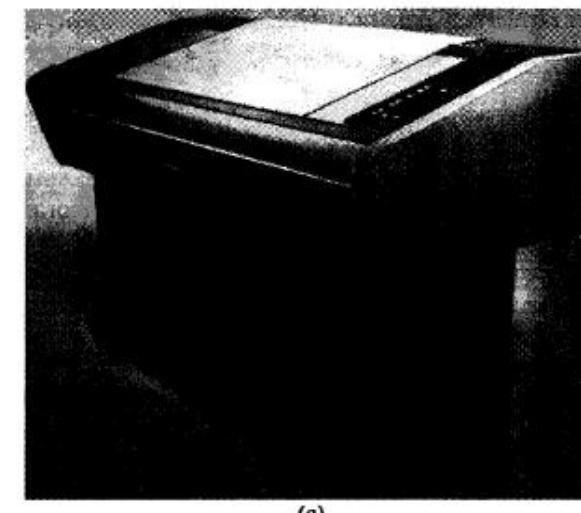


Рис. 1-45 Образцы размещения точек: a) без совмещения, b) совмещение по оси y , c) совмещение по осям x и y .

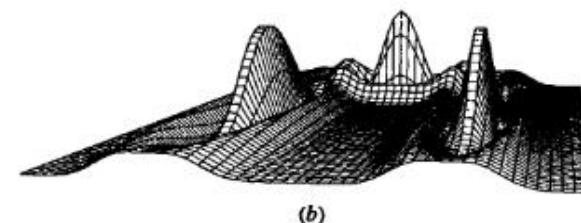
Типичная ширина электростатических графопостроителей достигает 72 дюймов, а длина изображения ограничивается только длиной рулона материала носителя (обычно это 500 футов). Разрешающая способность данных устройств зависит от количеством точек на дюйм (dpi). Точность определяется несколькими различными способами. Предполагается, что направление движения материала может быть изменено на противоположное, точность в этом случае определяется величиной ошибки, т. е. величиной расхождения в размещении двух точек, нанесенных одним и тем же пером при двух последовательных проходах.

Электростатические графопостроители являются растровыми устройствами вывода, поэтому для них имеет место лестничный эффект [1-1]. В устройствах с высоким разрешением данный эффект устраняется за счет совмещения отдельных точек. На рис. 1-45 показан типичный образец размещения точек. Обычный уровень совмещенности точек составляет от 30% до 50%. В связи с тем что электростатический графопостроитель является устройством с растровым сканированием, элементы изображения, например линии и многоугольники, должны иметь растровую развертку, а само изображение организуется по растровому принципу: сверху вниз, слева направо. Изображение может быть подготовлено на главном компьютере, однако необходимо соблюдать ряд требований организации вычислений. Кроме этого, для получения высокой скорости обработки картинки необходимо использовать быстрый интерфейс для передачи развернутого по раству изображения. Например, для графопостроителя шириной 36 дюймов, имеющего разрешение 400 точек на дюйм, требуется передавать данные от главного компьютера со скоростью 360 Кб/с для получения скорости рисования, равной 1 дюйм/с. С целью повышения быстродействия многие электростатические графопостроители работают со специальным контроллером [1-6]. На рис. 1-46 изображен типичный электростатический графопостроитель.

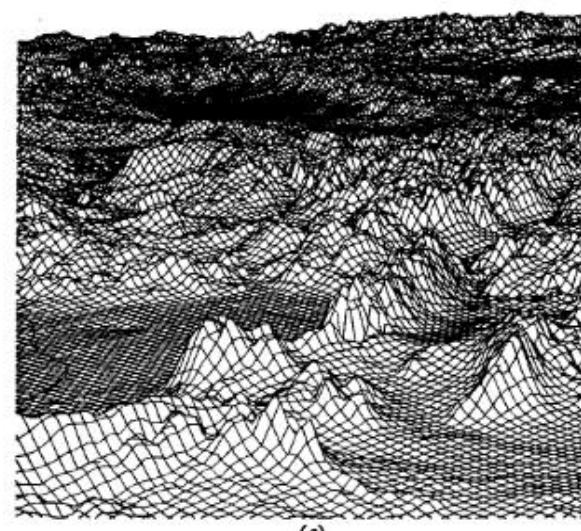
Существуют также цветные графопостроители, в которых используются три основных субтрактивных цвета [1-1] — голубой (cyan), пурпурный (magenta) и желтый (yellow) — и четвертый, черный, которые наносятся на материал за один или четыре прохода. В случае одного прохода, отдельные ряды перьев используются для нанесения каждого из цветов. При нескольких проходах каждый цвет наносится в течение одного прохода. Обычный графопостроитель поддерживает 7 цветов. Цветовая закраска сплошных многоугольников осуществляется путем псевдотонирования или заполнения узором [1-1]. Количество возможных цветов зависит от размеров используемого узора (например, 2×2 , 2×3 или 4×4) и от разрешения графопостроителя. Обычно при высоком разрешении графопостроитель способен выводить несколько тысяч цветов. Стандартная



(a)



(b)



(c)

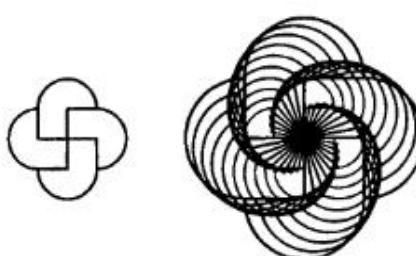
Рис. 1-46 a) Электростатический графопостроитель. (С разрешения фирмы Versatec Corp.) b) Низкое разрешение 100 dpi; c) Высокое разрешение 400 dpi.

ширина электростатических графопостроителей — 42 дюйма, а разрешение — 400 точек на дюйм.

1-15 ЛАЗЕРНЫЕ ПЕЧАТАЮЩИЕ УСТРОЙСТВА

Несмотря на то что лазерные печатающие устройства (принтеры) первоначально создавались для использования в типографском деле, они нашли достаточно широкое применение и в машинной графике при решении задач вывода графических изображений. Возможности этих устройств оказались особенно полезными в различного рода издательских системах, где возникает необходимость в компоновке на одной странице текстов и изображений [1-7]. Такие системы позволяют комбинировать выведенные различными шрифтами тексты, художественные иллюстрации и полутонаовые изображения, как, например, представлено на рис. 1-47.

Лазерный принтер принадлежит к семейству растровых устройств. Принципы его действия, схематично представленные на рис. 1-48, во многом подобны принципам, применяемым при ксерокопировании. Печатающий механизм имеет покрытый фоточувствительным веществом барабан, по поверхности которого может перемещаться луч лазера (газоразрядного или полученного от полупроводникового диода). При вращении барабана его покрытие получает электрический заряд, который сохраняется до момента падения на него когерентного лазерного луча. В результате прохождения луча на барабане появляются зоны (точки), не имеющие заряда, которые в совокупности образуют негативное изображение будущей картинки. Позиционирование луча на барабане осуществляется с помощью вращающегося многогранного отражателя. Черный порошок на пластиковой основе (тонер) притягивается к заряженным точкам на барабане, изображение с которого переносится затем на лист положительно заряженной бумаги. На завершающей стадии формирования изображения происходит фиксация тонера путем его всплавления под действием температуры и давления.



Картинка наверху слева формируется командами `gline()` и `arc()`. Затем путем масштабирования и поворота можно получить различные изображения

Рис. 1-47 Лазерный принтер позволяет формировать страницу путем комбинации текстов, иллюстраций и полутонаовых изображений. (С разрешения фирмы QMS, Inc.)



Рис. 1-48 Схема устройства лазерного принтера.

Обычно лазерный принтер имеет разрешение от 240×240 до 600×600 т/дюйм. Наиболее типичное разрешение — это 300×300 т/дюйм. Лабораторные эксперименты показывают, что практически высший предел возможностей данной технологии — 800×1000 т/дюйм. При более высоком разрешении происходит спекание соседних точек тонера. Несмотря на то что лазерные принтеры используются для типографского набора, их обычное разрешение значительно меньше, чем разрешение, полученное при хорошем фотонаборе ($1200-5000$ т/дюйм). Качество современных принтеров ограничено размерами отдельной точки, однако лабораторные эксперименты по изменению размеров точки дают обнадеживающие результаты. Изменение размеров точки позволяет увеличить гибкость при формировании полутонаовых изображений [1-1]. Хотя размеры лазерных принтеров определяют формат используемой для них бумаги, в настоящее время выпускаются устройства шириной 36 дюймов с разрешением 400 т/дюйм, которые позволяют получать изображения форматов D (22 × 44) и E (34 × 44) — формат А0 по ЕСКД.

Лазерная печать является достаточно продолжительным процессом. Вывод страницы происходит с постоянной скоростью, так как для получения высокого разрешения требуется и высокая скорость обработки. Например, при разрешении 300×300 для печати со скоростью 8 страниц в минуту при размере $8,5 \times 11$ дюймов, с печатным полем $7,5 \times 10,5$ обеспечение непрерывной бесперебойной работы потребует скорости 1 Мбит/с. Для поддержания такой же высокой скорости, как и в электростатических графопостроителях, используется специальный контроллер, общая схема которого приведена на рис. 1-49. Главный компьютер выдает описание страницы на языке высокого уровня [1-7]. Видеоконтроллер/процессор отвечает за генерацию шрифтов, линий или полутонаовых изображений и за передачу всего необходимого потока битовой информации (точек) на печатающее устройство.

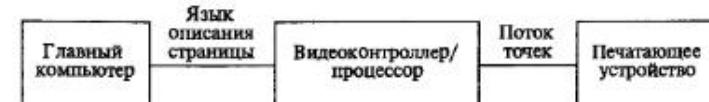


Рис. 1-49 Конфигурация вычислительной системы лазерного принтера.

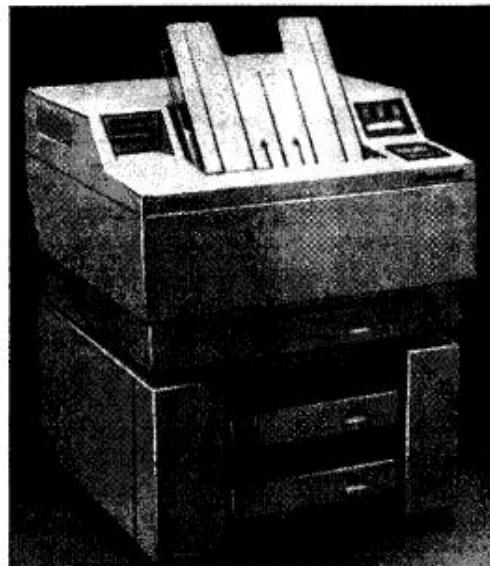


Рис. 1-50 Лазерный принтер. (С разрешения фирмы Toshiba America, Inc.)

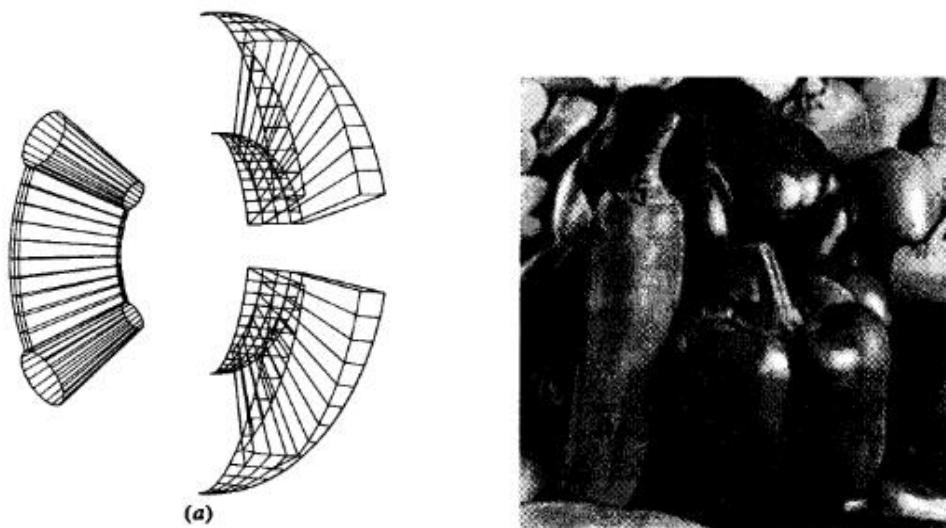


Рис. 1-51 Примеры изображений, выводимых лазерным принтером: а) векторная графика; б) полутоновые поля.

На рис. 1-50 показан типичный лазерный принтер, а на рис. 1-51 приведены примеры обычных векторных и полутоновых изображений.

1-16 МАТРИЧНЫЕ ГРАФОПОСТРОИТЕЛИ

Принтеры использовались еще на заре становления машинной графики, когда применялись обычные строковые устройства, принадлежащие к типу устройств с растровым сканированием. С помощью таких принтеров можно получать изображения, графики и даже полутоновые картинки. Вполне естественно, что при разрешении 1/6 дюйма по вертикали и 0.1 по горизонтали изображения были довольно грубыми.

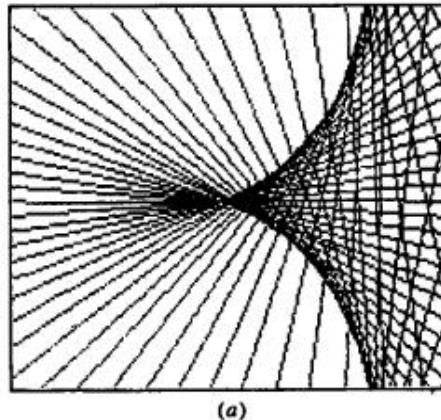
В машинной графике применяется множество принтеров, основанных на использовании различных методов и принципов печати: ленточные или построчные устройства, барабанные принтеры, цепные печатающие устройства и лепестковые принтеры. Однако наибольшую популярность и распространение получили именно матричные устройства.

Печатающий механизм в матричных принтерах представляет собой матрицу из тонких перьев или иголок, закрепленную на печатающей головке, которая перемещается вдоль страницы. Основу механизма составляет электромагнит для одной или нескольких стальных пружин, смонтированных на иглах или молоточках. В состоянии покоя молоточки удерживаются постоянным магнитом и не соприкасаются с красящей лентой. Для нейтрализации действия этого постоянного магнита используется электромагнит. При его включении стальные пружины подают молоточки вперед, вызывая соприкосновение иголки с лентой. В результате на бумагу, которая расположена между лентой и опорой, наносится точка. После этого электромагнит отключается, молоточек возвращается в первоначальное положение. В конструкциях со свободными молоточками их возврат и освобождение происходит в состоянии возбуждения электромагнита.

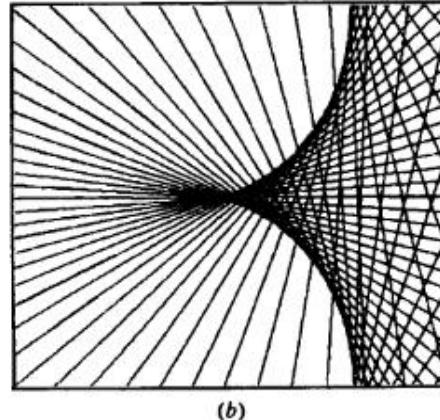
Существует множество различных конфигураций расположения иголок (рис. 1-52). При невысокой плотности размещения иголок размером от 0.012 до 0.013 дюйма получается изображение точки размером 0.014–0.015 дюймов, что позволяет иметь разрешение от 60 до 80 т/дюйм. При большей плотности размещения иголок размером 0.008 дюймов образуется точка размером 0.009–0.010 дюйма, а разрешение составляет 180 т/дюйм. Таким образом, видно, что при высокой плотности точек получается отпечаток с перекрытием. В зависи-



Рис. 1-52 Образцы размещения иголок в матричном печатающем устройстве.



(a)



(b)

Рис. 1-53 Сравнение качества вывода матричных печатающих устройств: а) 9 иголок — разрешение 70 точек на дюйм; б) 24 иголки — разрешение 180 точек.

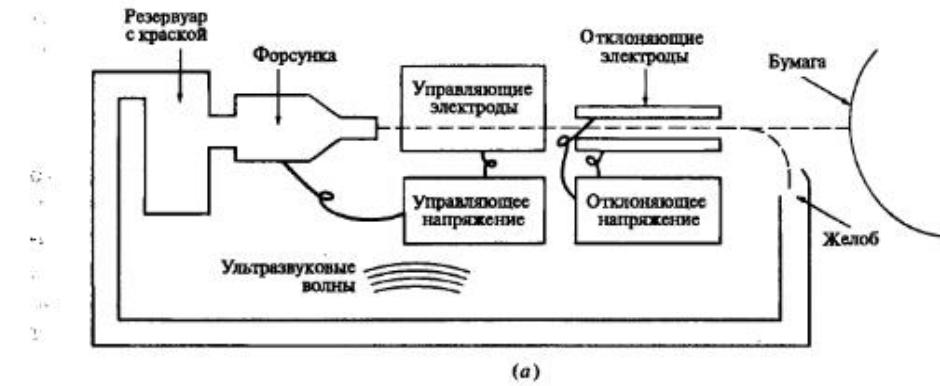
сти от плотности точек перекрытие может составить от 20 до 50%. Перекрытие точек уменьшает уровень смещения или ступенчатый эффект [1-1], что способствует улучшению качества изображения. На рис. 1-53 представлена картинка, полученная на матричных принтерах с разрешением 70 и 180 т/дюйм и иллюстрирующая данный эффект.

Цветное изображение на матричных печатающих устройствах получается с помощью многоцветных красящих лент. Так как бумага обладает средней отражающей способностью, то используется четырехцветная лента, на которой имеются три основных цвета: голубой, пурпурный, желтый (CMY), а также черный цвет. Использование методов псевдотонирования и заполнения узорами [1-1, 1-8] позволяет получать оттенки разного цвета.

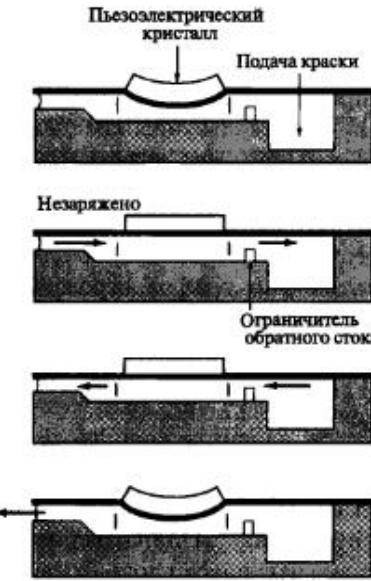
1-17 УСТРОЙСТВО СТРУЙНОЙ ПЕЧАТИ

Устройства струйной печати являются растровыми устройствами и используются главным образом в недорогих системах формирования цветных изображений. Основной принцип струйной печати состоит в распылении микроскопических капель краски на носитель изображения. Существуют два типа устройств струйной печати: с непрерывным распылением и с распылением по запросу.

Устройства первого типа формируют поток частиц путем распыления краски через форсунку. Такой поток прерывается с помощью ультразвуковых волн. Капельки краски, которые должны быть нанесены на бумагу, получают электрический заряд. Специальные отражательные пластины управляют направлением движения частиц. Все незаряженные частицы отклоняются таким образом, что они попадают в желоб и возвращаются затем в резервуар. Обычно в качестве носителя изображения используется бумага или прозрачная пленка. Типичная система струйной печати показана на рис. 1-54а. В случае применения технологии распыления по запросу, как следует из ее названия, выброс краски происходит, только если требуется закрасить определенную точку. Краска поступает из ре-



(a)



(b)

Рис. 1-54 Принципиальная схема технологии струйной печати: а) непрерывное распыление; б) распыление по запросу.

зервуара и попадает под давлением в форсунку. Далее с помощью пьезоэлектрического кристалла, на который подается напряжение, происходит вспрыскивание краски через голову форсунки. Кристалл расширяется под действием напряжения, уменьшая объем камеры с краской, что и вызывает ее выброс из форсунки. При отключении напряжения кристалл сжимается, увеличивая объем резервуара, и это приводит к всасыванию краски обратно в форсунку. На рис. 1-54б представлена иллюстрация данного процесса.

Разрешающая способность устройств струйной печати определяется размерами капли краски, а следовательно, и форсунки. При использовании максимально узкой форсунки, возникает ряд проблем с обеспечением свободного прохожде-

ния краски, предотвращением загрязнения форсунки, а также с удалением пузырьков воздуха. Обычное разрешение устройств данного типа составляет 120–400 т/дюйм. Лабораторные эксперименты показывают, что возможно разрешение в 1000 т/дюйм.

Цветные устройства струйной печати имеют, как правило, четыре форсунки, три для основных цветов — голубого, пурпурного, желтого — и одну для черного цвета. Основное преимущество струйной технологии состоит в способности смешивать цвета. Так как жидкие частицы краски наносятся за один проход носителя, они успевают перемешаться еще до высыхания. Это позволяет получать глубину и резкость цвета, не достижимую при использовании других технологий. Применение методов псевдотонирования и узоров, как уже отмечалось, позволяет получать несколько тысяч цветов [1-1, 1-8].

1-18 УСТРОЙСТВО ТЕРМОПЕЧАТИ

Различают два основных принципа термопечати: с прямым и косвенным воздействием тепла. В первом случае используется специальная термоочувствительная бумага, которая изменяет свой цвет под действием тепла. Изображение на такой бумаге формируется печатающим узлом, который воздействует на бумагу в процессе движения. Устройство пишущего узла матричного принтера и печатающие головки в виде полного ряда пишущих узлов аналогичны устройству узлов, используемых в электростатических графопостроителях. Устройства термопечати позволяют получать разрешение в 40 т/дюйм. Главные и наиболее серьезные ограничения технологии с прямой передачей тепла связаны с необходимостью применения специальной бумаги, которая имеет свойство выцветать с течением времени.

В технологии с косвенным действием тепла используется тонкая пленка или бумажная лента, покрытая чернилами на восковой основе. Тепловыделяющие элементы, обычно располагаемые позади ленты, расплавляют восковое покрытие, и пигменты чернил попадают на бумагу. Однаково часто употребляются матричные печатающие узлы, перемещающиеся вдоль бумаги, и печатающие головки в виде полного ряда пишущих узлов. На рис. 1-55 представлена схема технологии

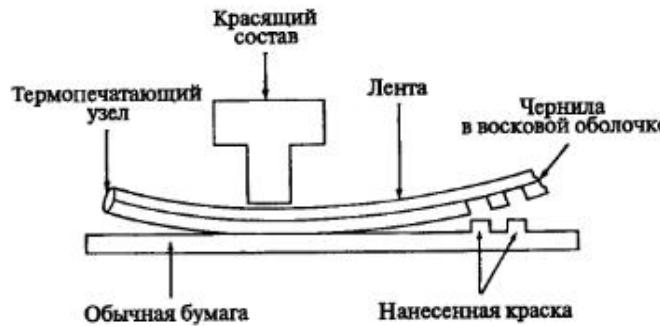


Рис. 1-55 Технология с косвенным тепловым воздействием.

с косвенным воздействием тепла. В качестве основы для нанесения изображения применяется обычная бумага. Цветные картинки получаются путем использования многоцветных лент. Традиционные семь цветов возникают в результате комбинации трех основных цветов (CMY) и черного цвета. Для формирования расширенной цветовой палитры применяются шаблоны и штриховка [1-1, 1-8]. Так как чернила почти не впитываются в бумагу, то можно получать блестящие цвета. Обычная разрешающая способность для устройств данного типа составляет от 100 до 400 т/дюйм. Чернила, употребляемые при технологии с косвенным воздействием, отличаются долговечностью.

1-19 ПЕРЬЕВЫЕ ГРАФОПОСТРОИТЕЛИ.

Перьевые графопостроители (прежде всего цифровые) делятся на три вида: планшетные, барабанные и графопостроители с нажимным роликом. На рис. 1-56 даны принципиальные схемы каждого из этих устройств.

В планшетном графопостроителе с подвижным кронштейном (рис. 1-56а) носитель изображения закрепляется на ложе графопостроителя. Перемещение пишущего узла на плоскости осуществляется при движении кронштейна по ширине ложа для оси x и при движении собственно узла по кронштейну для оси y . На

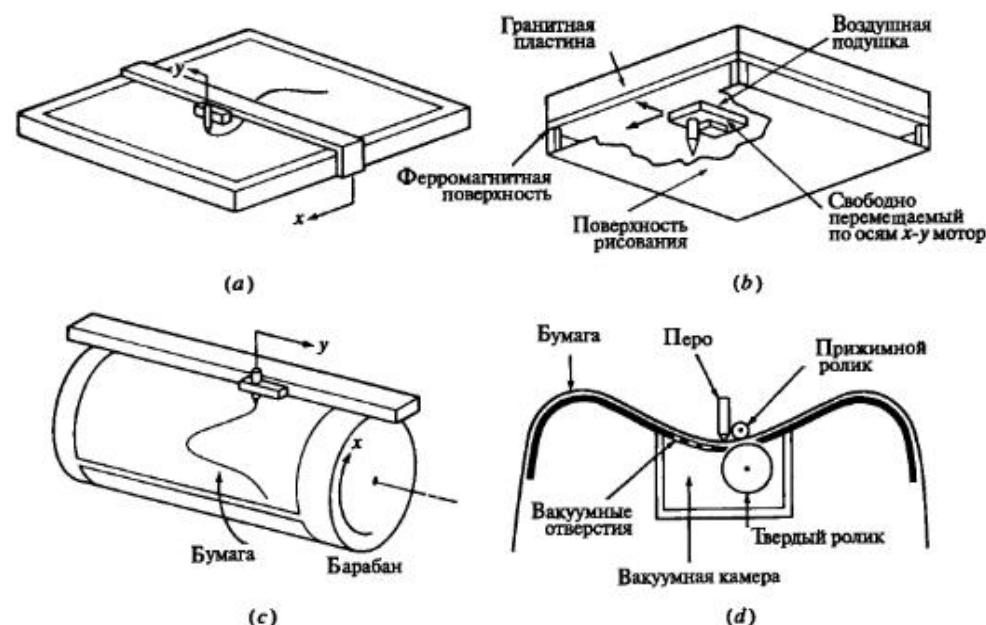


Рис. 1-56 Схемы действия перьевых графопостроителей разных типов: а) планшетный с подвижным кронштейном; б) планшетный с подвижным пишущим узлом; в) барабанный; г) с нажимным роликом.



Рис. 1-57 Планшетный графопостроитель с подвижным кронштейном. (С разрешения фирмы Xynetics, Inc.)



Рис. 1-58 Планшетный графопостроитель с подвижным пишущим узлом.

рис. 1-57 приведен типичный планшетный графопостроитель с подвижным кронштейном.

В планшетном графопостроителе с подвижным пишущим узлом (рис. 1-56б) применяется устройство рисования, которое, находясь в подвешенном состоянии, перемещается под действием магнитных сил. Данный принцип обеспечивает почти полное отсутствие трения. Управление движением пишущего узла по двум направлениям осуществляется на основе электромагнитного воздействия по принципу сервомотора. Пример графопостроителя данного типа представлен на рис. 1-58.

Крепление носителя изображения на ложе крупноформатных планшетных графопостроителей осуществляется с помощью вакуума. Данные устройства имеют разрешение не менее 0.004 дюйма и скорость 3600 дюйм/мин. Размеры планшетных графопостроителей колеблются от 8.5 × 11 дюймов до нескольких десятков футов в ширину и длину. В качестве носителя изображения может ис-

пользоваться множество материалов, таких, как бумага, пергамент, фотопленка, копировальная поверхность, лист металла и ткани. Этим разнообразным типам носителей соответствует множество устройств рисования, среди которых шариковая ручка, фломастеры и обычные чернильные перья, световые и лазерные лучи, инструменты для гравировки, копировальные резцы и резцы по ткани. Такое разнообразие типов носителей и пишущих узлов предполагает широкий спектр различных сфер применения планшетных графопостроителей. Кроме их традиционного использования для рисования, крупноформатные графопостроители применяются для создания шаблонов печатных плат и интегральных схем. Носителем изображения в этом случае является фотопленка, а пишущим инструментом луч света. Различные модификации планшетных графопостроителей используются также и для раскроек мужской и женской одежды и даже парусов. В качестве носителя здесь выступает соответствующая ткань, а пишущим узлом является нож либо лазерный луч.

Более дешевые и небольшие планшетные графопостроители не обладают такой гибкостью применения и имеют более низкую скорость и разрешающую способность. Обычно в таких устройствах используется принцип электростатического управления носителем, в качестве которого, как правило, применяется бумага, пергамент или прозрачная пленка.

Барабанные графопостроители (рис. 1-56с) являются более сложными в механическом плане устройствами, чем планшетные. Пишущий узел в них перемещается вдоль неподвижного кронштейна в одном направлении. Сам носитель вращается вперед и назад под кронштейном, обеспечивая таким образом отрисовку по другому направлению.

Носитель закрепляется на барабане при помощи вакуума либо прижимных роликов. Тип материала носителя определяется его свойствами легко вращаться. Обычно это бумага или пергамент определенной ширины. Длина носителя ограничена только длиной ленты в рулоне. Несмотря на то что разрешающая способность графопостроителей данного типа примерно равна разрешению планшетных устройств (0.0005 дюйма), скорость рисования в них, как правило, значительно ниже.

Графопостроители с прижимным роликом представляют собой смесь планшетных и барабанных устройств. Носитель изображения размещается в них между прижимным роликом и цилиндрической поверхностью рисования (рис. 1-56д). При вращении прижимного ролика носитель движется вперед и назад под неподвижным кронштейном, по которому, в свою очередь, перемещается пишущий узел. В этих графопостроителях применяется накопитель для отдельных листов либо для рулона, что и определяет формат бумаги, пергамента или прозрачной пленки. Обычная разрешающая способность данных устройств составляет 0.001 дюйм, а скорость равна 180 дюйм/мин. На рис. 1-59 показан графопостроитель с прижимным роликом.

Большинство перьевых графопостроителей работают в инкрементальном режиме и приводятся в движение шаговым мотором. Пишущий узел в них перемещается над поверхностью рисования по восьми возможным направлениям (рис. 1-60) за некоторое число небольших шагов. Результатирующая линия, проведенная в направлении, отличном от указанных на рис. 1-60, будет иметь ступенчатый ха-

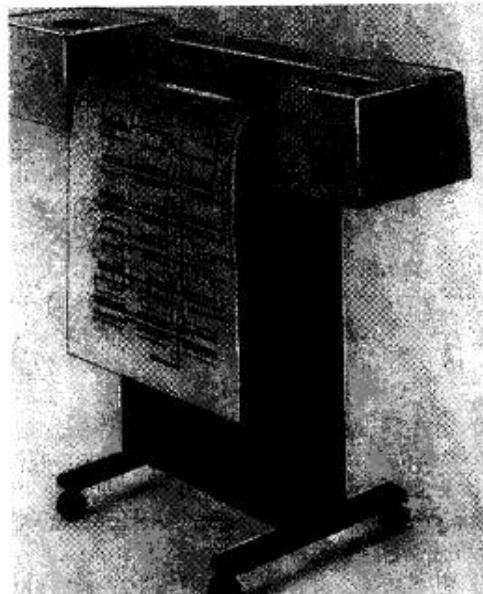


Рис. 1-59 Графопостроитель с прижимным роликом. (С разрешения фирмы Hewlett-Packard. Co.)

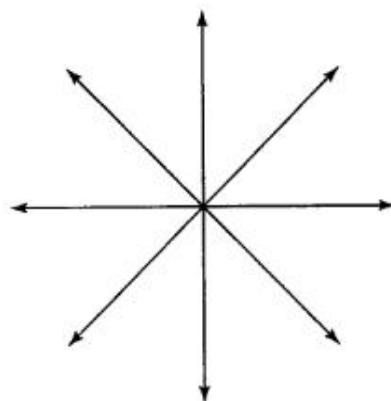


Рис. 1-60 Направления движения в инкрементальных графопостроителях.

рактер. Появление такой «лестницы», или зазубренной линии, является следствием ключевого явления, получившего название лестничный эффект (алиазинг), т. е. дефект изображения линии вследствие дискретности раstra [1-1]. Из всех алгоритмов построения линий, имеющих наилучшие показатели оптимальности выбора из восьми возможных направлений, алгоритм Брезенхема [1-1] является наиболее удачным. Данный алгоритм, часто ассоциируемый с растровыми устройствами, первоначально был разработан именно для цифровых инкрементальных графопостроителей.



Рис. 1-61 Пример типичного изображения, получаемого на цифровом инкрементальном графопостроителе.

Корректность направления движения достигается путем использования серводвигателей, которые управляют пишущим узлом графопостроителя. В этом случае обратная связь в цепи постоянно корректирует одновременно оба мотора для получения гладкой линии.

Скорость является важной, но не единственной существенной характеристикой графопостроителя. В некоторых случаях ускорение оказывает гораздо большее влияние на общее быстродействие. Дело в том, что чертеж состоит из коротких сегментов линий и поэтому графопостроитель должен уметь стартовать и останавливаться между сегментами. Большое ускорение позволяет графопостроителю быстрее достигать нужной скорости. Возможная в настоящее время величина ускорения превышает 4 g. Многие контроллеры графопостроителей комплектуются также специальными программами, позволяющими заранее избегать лишних остановок, например, в случае, когда для представления кривой используется последовательность соединенных друг с другом коротких отрезков.

В отличие от многих графических устройств, первые графопостроители позволяют получать достаточно высококачественные изображения, такие как на рис. 1-61.

1-20 ЦВЕТНЫЕ ВИДЕОКАМЕРЫ

С появлением высококачественных цветных растровых дисплеев в сочетании с цветными буферами кадров (§ 1-9), способных выводить свыше 16 млн. цветов, возникла необходимость в способе представления полученных результатов. В настоящее время существует единственная такая технология, основанная на использовании цветной фотопленки. Цветная видеокамера — это наиболее часто используемое в машинной графике устройство.

Совсем простая принципиальная схема цветной камеры приведена на рис. 1-62. В методе, называемом записью последовательных полей, используется небольшая, диаметром 5 дюймов, монохроматическая ЭЛТ с идеально плоским

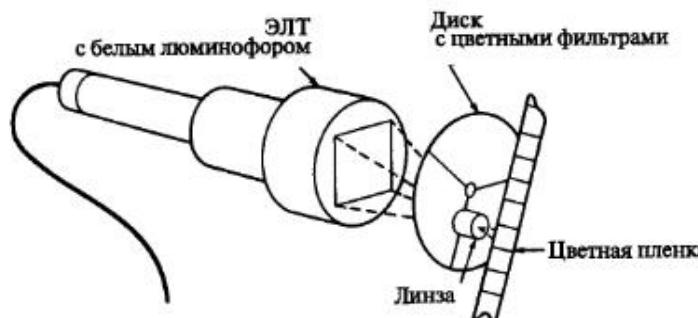


Рис. 1-62 Принципиальная схема цветной камеры.

экраном. На оптической оси от камеры к ЭЛТ сначала помещается красный фонофильр, пропускающий только красные лучи от ЭЛТ, которые и попадают на пленку. После этого экран ЭЛТ гаснет, красный фильтр заменяется на зеленый, открывающий путь только зеленым лучам. Пленка повторно экспонируется, оставаясь в прежнем положении. В заключение процесс повторяется для синей компоненты изображения. Такая трижды экспонированная пленка содержит за-конченное изображение рисунка.

Успех всего процесса определяется качеством ЭЛТ [1-1], для уменьшения в ней искажений применяется плоский, а не изогнутый экран, устанавливаемый на большинство мониторов. Стекло ЭЛТ также должно быть высокого качества, без пузырьков, царапин и других дефектов. Для того чтобы обеспечить примерно одинаковую реакцию на все три сигнала (красный, зеленый, синий), применяется широкополосный люминофор, слой которого должен быть без пятен и узоров, иметь прекрасную зернистость и однородную текстуру. Размеры точки электронного луча должны быть минимальными размерами для достижения повышенной разрешающей способности, а электромагнитная фокусирующая система должна быть отличного качества.

При условии аккуратной эксплуатации разрешение такого устройства составляет 3000×5000 т/дюйм при диаметре монитора 5 дюймов. Отметим, что разрешение и адресуемость — это не идентичные понятия. Разрешением измеряется количество отдельных различимых точек, в то же время адресуемость указывает число позиций, в которые может быть помещен электронный луч. Из-за ограниченности возможностей человеческого глаза картинка может иметь адресуемость выше порога, за которым глаз уже не различает отдельных точек изображения.

1-21 АКТИВНЫЕ И ПАССИВНЫЕ ГРАФИЧЕСКИЕ УСТРОЙСТВА

Разберем различия между пассивными и активными графическими устройствами. Пассивные устройства просто формируют под управлением компьютера изображения, т. е. они позволяют обеспечивать графический интерфейс между компьютером и пользователем. Матричные принтеры, электростатические и перьевые графопостроители, устройства записи на пленку, дисплеи с запоминающей трубкой, а также мониторы с растровой ЭЛТ — все это примеры пассивных гра-

фических устройств. Некоторые образцы этих устройств вместе с типичными картинками, которые они способны формировать, приведены на рис. 1-17, 1-22, 1-37, 1-40, 1-46, 1-50, 1-51, 1-53, 1-57, 1-58, 1-59 и 1-61.

Активные графические устройства позволяют пользователю взаимодействовать с компьютером при помощи графики. Обычно это означает, что пользователь получает координатную информацию в некотором преобразованном виде, т. е. в виде, отличном от явного вывода соответствующих чисел. Так как картинки, кривые или поверхности можно рассматривать в виде матрицы координат, то пользователь получает достоверную иллюстративную информацию. Обычно активные устройства обладают возможностью позиционирования курсора и считывания его нового положения. Типичные активные графические устройства включают алфавитно-цифровые клавиатуры (рис. 1-6), функциональные кнопки (рис. 1-13), координатные колеса, наборы ручек для ввода скалярных значений (рис. 1-12), оцифровывающий или аналоговый планшеты (рис. 1-7, 1-8), световые перья (рис. 1-15), управляющие рычаги (рис. 1-9), координатный шар (рис. 1-10) или мышь (рис. 1-11). Несмотря на то что эти устройства могут использоваться самостоятельно, обычно они требуют применения каких-либо пассивных устройств. Как правило это ЭЛТ.

1-22 ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ МАШИННОЙ ГРАФИКИ

Весь комплекс программного обеспечения машинной графики не только реализует идеи, затронутые в данной книге и в работе [1-1], но и воплощает алгоритмы организации структур данных, проектирования и управления базами данных, идеологию и требования к человеко-машинному интерфейсу, охватывает языки программирования и операционные системы. Исследование всех этих вопросов выходит за рамки данной книги, однако они затрагиваются при обсуждении вопросов проектирования систем машинной графики.

В настоящее время разработаны и продолжают создаваться многочисленные графические стандарты, которые в общем случае можно объединить в три группы. Первая группа — это графический прикладной интерфейс, преобразующий идеи в форму, понятную компьютеру. Для этих целей применяются стандарты представления данных: графическая корневая система (GKS), иерархический интерактивный графический стандарт программиста (PHIGS, PHIGS+) (приложение А).

Вторая группа относится к хранению и передаче данных между графическими системами и между САПР и системами АСТПП. В данной области имеется стандарт IGES (формат передачи графических данных).

В третью группу входит стандарт для определения интерфейса между программными системами машинной графики или наборами данных и системой передачи, а также стандарт определения требований к аппаратно зависимым драйверам графических устройств. В данной области существуют стандарты: машинно-графический метафайл (CGM) и виртуальный интерфейс устройств (VDI).

Все эти стандарты имеют очень важное значение и представляют собой попытку обеспечения совместимости программ, пользователей и данных. В приложении А эти стандарты рассматриваются более развернуто, но не очень детально.

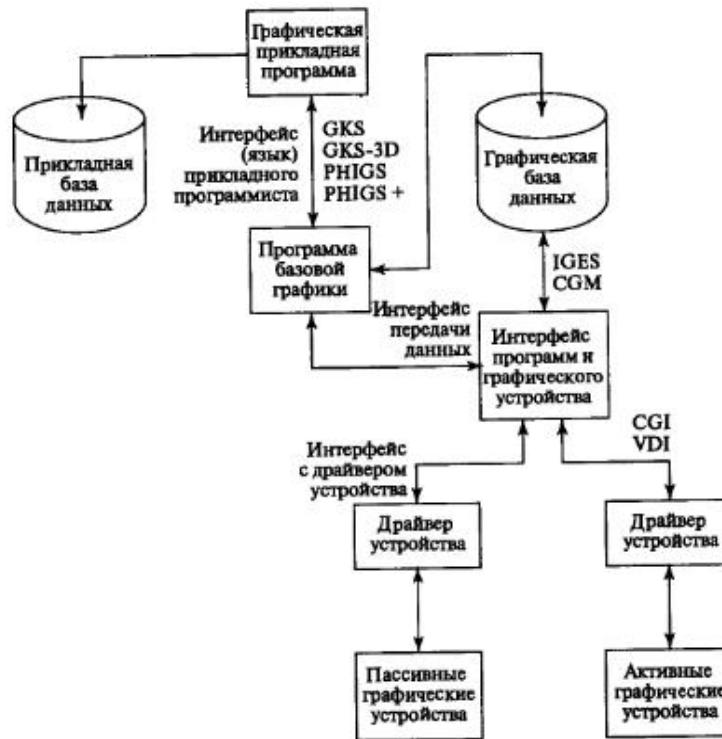


Рис. 1-63 Структура программного обеспечения машинной графики.

На рис. 1-63 показаны основные уровни организации программного обеспечения машинной графики и место каждого из стандартов. Однако задача этой книги и работы [1-1] состоит не в обсуждении стандартов, а в исследовании основополагающих концепций, математических и процедурных элементов машинной графики. Поэтому программам машинной графики уделяется минимальное внимание на уровне основных понятий. Программные системы машинной графики, реализующие все эти концепции, представлены в приложениях А и Г. На основе всех изложенных и разбираемых впоследствии концепций из этой книги и из работы [1-1], а также концепций организации структур данных и проектирования баз данных можно разработать и реализовать функционально полную программную систему машинной графики. Эти концепции помогут также разобраться в любой из существующих систем машинной графики.

1-23 ЛИТЕРАТУРА

- 1-1 Rogers D. F. *Procedural Elements for Computer Graphics*, McGraw-Hill, Book Co., New York, 1985. (Имеется русский перевод: Роджерс Д. Алгоритмические основы машинной графики. — М.: Мир, 1989.)
- 1-2 Foley J. D., Wallace V. L. *The Art of Natural Man-Machine Conversation*, Proc. IEEE, Vol. 62, pp. 462–471, 1974.

- 1-3 Wallace V. L. *The Semantics of Graphic Input Devices*, *Comp. Graph.*, Vol. 10, pp. 61–65, 1976.
- 1-4 Olson M. *System Design Considerations for Graphics Input Devices*, *Computer*, pp. 9–18, Nov. 1978.
- 1-5 Davis M. R., Ellis T. O. *The RAND Tablet: A Man-Machine Graphical Communication Device*, *AFIPS Conf. Proc.*, Vol. 26, Part I, 1964 FJCC, pp. 325–332, 1964.
- 1-6 Ben-Dor A., Jones B. *New Graphics Controller for Electrostatic Plotting*, *IEEE Comp. Graph. & Appl.*, Vol. 6, No. 1, pp. 16–25, 1986.
- 1-7 Adobe Systems Inc., *Postscript Language Reference Manual*, Addison-Wesley Publishing Co., Reading, MA, 1985.
- 1-8 Kubo S. *Continuous Color Presentation Using a Low-Cost Ink Jet Printer*, *Proc. Comp. Graph. Tokyo*, 84, 24–27. Apr. 1984, Tokyo, Japan, T3-6, pp. 1–10, 1984.
- 1-9 Constantine J. M., Jr., *Electron Optical Technique for an Ultra High Resolution Color Film Recorder*, *Journal of Photooptical Instrument Engineers*, 1984.

ДВУМЕРНЫЕ ПРЕОБРАЗОВАНИЯ

2-1 ВВЕДЕНИЕ

Изучение математического аппарата, лежащего в основе машинной графики, мы начнем с рассмотрения способов вывода и преобразования точек и линий. Эти способы наряду с соответствующими алгоритмами рисования используются для изображения объектов или визуализации графической информации. Возможность проводить преобразования точек и линий является фундаментом машинной графики. Нарисованный объект может быть представлен в нужном масштабе, повернут, перемещен, преобразован или модифицирован в соответствии с требованиями решаемой задачи. Все эти манипуляции с изображением можно выполнить, используя математический аппарат, изложенный в данной и последующих главах.

2-2 ИЗОБРАЖЕНИЕ ТОЧЕК

Точка представляется на плоскости двумя своими координатами, которые определяются как элементы матрицы размером 1×2 $[x \ y]$. В трехмерном пространстве используется матрица размером 1×3 $[x \ y \ z]$.

Иначе говоря, точка может задаваться в виде вектор-столбца $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ в двумерном пространстве или в виде

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

в трехмерном. Строку $[x \ y]$ или столбец $\begin{bmatrix} x \\ y \end{bmatrix}$ часто называют координатным вектором. В этой книге для формирования такого вектора используется матрица-строка, т. е. множество точек, каждая из которых определяет координатный вектор в некоторой системе измерения. Данное множество хранится в компьютере в виде матрицы или массива чисел. Положением точек можно управлять путем манипулирования соответствующей матрицей. Линии, соединяющие точки, формируют отрезки, кривые и картинки.

2-3 ПРЕОБРАЗОВАНИЯ И МАТРИЦЫ

В качестве элементов матрицы могут фигурировать различные величины: числа, сетки или коэффициенты системы уравнений. Правила в матричной алгебре определяют допустимые операции над элементами (приложение B). Многие физические задачи удобно выражаются в матричном представлении. Для моделей физических систем задача обычно ставится следующим образом: даны матрицы $[A]$ и $[B]$, найти результирующую матрицу $[T]$, такую, что $[A][T] = [B]$. В этом случае решением является матрица $[T] = [A]^{-1}[B]$, где $[A]^{-1}$ — матрица, обратная к квадратной матрице $[A]$.

В то же время матрицу $[T]$ можно интерпретировать как геометрический оператор. В этом случае для выполнения геометрического преобразования точек, представленных векторами положений в матрице $[A]$, используется умножение матриц. Предположим, что матрицы $[A]$ и $[T]$ известны. Требуется определить элементы матрицы $[B]$. Представление $[T]$ как геометрического оператора является основой математических преобразований, используемых в машинной графике.

2-4 ПРЕОБРАЗОВАНИЕ ТОЧЕК

Рассмотрим результаты умножения матрицы $[x \ y]$, содержащей координаты точки P , на матрицу общего преобразования размером 2×2 :

$$[X][T] = [x \ y] \begin{bmatrix} a & b \\ c & d \end{bmatrix} = [(ax + cy)(bx + dy)] = [x^* \ y^*]. \quad (2-1)$$

Данная запись означает, что исходные координаты точки x и y преобразуются в x^* и y^* , где $x^* = ax + cy$, $y^* = bx + dy$ (приложение B). Представляют интерес значения x^* , y^* — координаты результирующей, преобразованной точки P . Рассмотрим некоторые специальные случаи.

При $a = d = 1$ и $c = b = 0$ преобразование сводится к единичной матрице

$$[X][T] = [x \ y] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [x \ y] = [x^* \ y^*], \quad (2-2)$$

и координаты точки P останутся неизменными. Как и следовало ожидать, в линейной алгебре умножение на единичную матрицу эквивалентно умножению на 1 в обычной алгебре.

В случае $d = 1$, $b = c = 0$

$$[X][T] = [x \ y] \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} = [ax \ y] = [x^* \ y^*], \quad (2-3)$$

где $x^* = ax$ — результат масштабирования координаты X . Эффект такого преобразования показан на рис. 2-1а. Рассмотрим теперь еще случай $b = c = 0$, т. е.

$$[X][T] = [x \ y] \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} = [ax \ dy] = [x^* \ y^*]. \quad (2-4)$$

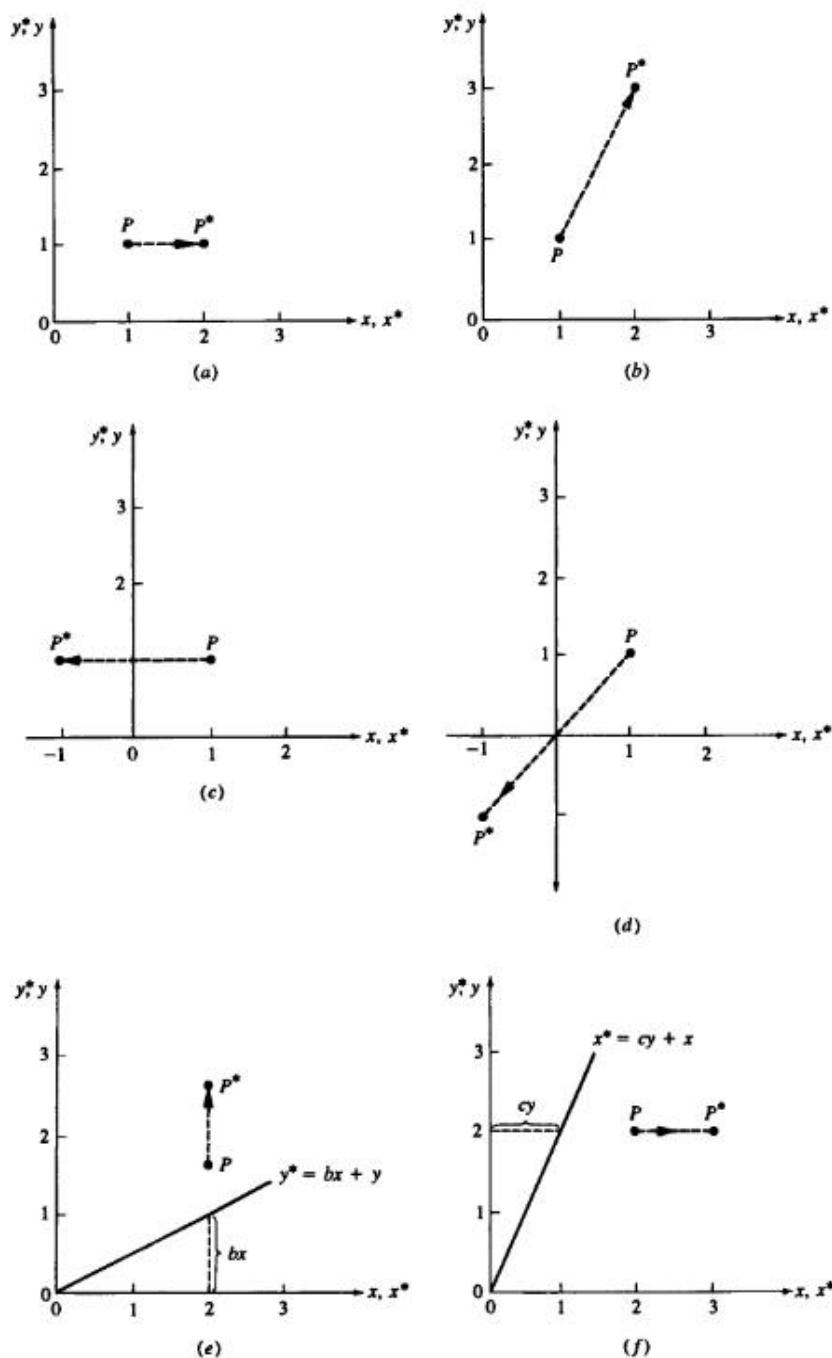


Рис. 2-1 Преобразование точек.

Данное преобразование вызывает изменение обеих координат x и y вектора P (рис. 2-1b). Если $a \neq d$, то координаты масштабируются различным образом. При $a = d > 1$ происходит растяжение вектора P или масштабирование координат. Если $0 < a = d < 1$, то имеет место сжатие.

Если значение a или d отрицательное, то вектор отражается относительно координатных осей или относительно плоскости. Чтобы убедиться в этом, возьмем $b = c = 0$, $d = 1$ и $a = 1$, тогда

$$[X][T] = [x \ y] \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = [-x \ y] = [x^* \ y^*], \quad (2-5)$$

и в результате получаем симметричное отражение относительно оси y (рис. 2-1c). Если $b = c = 0$, $a = 1$, $d = -1$, то выполняется симметричное отражение относительно оси x . Если $b = c = 0$, $a = d < 0$, то происходит отражение относительно начала координат, это показано на рис. 2-1d, где $a = -1$, $d = 1$. Заметим, что обе операции отражения и масштабирование зависят только от диагональных членов матрицы преобразования.

Рассмотрим теперь случай с недиагональными членами. Возьмем сначала значения $a = d = 1$, $c = 0$, тогда

$$[X][T] = [x \ y] \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} = [x \ (bx + y)] = [x^* \ y^*]. \quad (2-6)$$

Заметим, что координата x точки P осталась неизменной, тогда как координата y^* линейно зависит от исходных координат. Данное преобразование называется сдвигом (рис. 2-1e). Аналогично, в случае, когда $a = d = 1$, $b = 0$, преобразование приведет к сдвигу пропорционально координате y (рис. 2-1f). Таким образом, видно, что недиагональные члены матрицы преобразования создают эффект сдвига координат вектора точки P .

Прежде чем закончить с преобразованием точек, разберем действие общего преобразования, заданного выражением (2-1), когда начальный вектор лежит в точке начала координат, т. е.

$$[x \ y] \begin{bmatrix} a & b \\ c & d \end{bmatrix} = [(ax + cy) \ (bx + dy)] = [x^* \ y^*]$$

или, в случае начала координат,

$$[0 \ 0] \begin{bmatrix} a & b \\ c & d \end{bmatrix} = [0 \ 0] = [x^* \ y^*].$$

Видно, что начало координат инвариантно относительно преобразования общего вида. Это ограничение устраняется при использовании однородных координат.

2-5 ПРЕОБРАЗОВАНИЕ ПРЯМЫХ ЛИНИЙ

Прямую линию можно определить с помощью двух векторов, задающих координаты ее конечных точек. Расположение и направление линии, соединяющей две

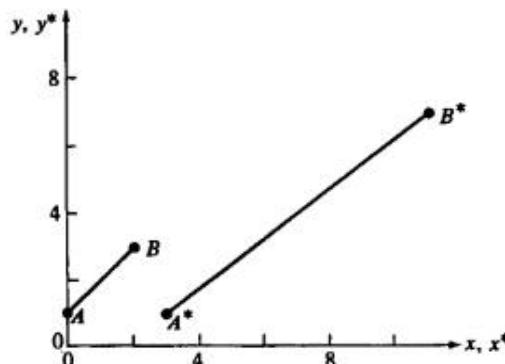


Рис. 2-2 Преобразование отрезков.

эти точки, может изменяться в зависимости от положений векторов. Реальный вид изображения линии зависит от типа используемого дисплея. В этом разделе мы рассмотрим только математические операции над конечными точками линии.

На рис. 2-2 изображена прямая линия, проходящая между двумя точками A и B . Положение векторов точек A и B задается следующим образом: $[A] = [0 \ 1]$ и $[B] = [2 \ 3]$.

Рассмотрим матрицу преобразования

$$T = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}, \quad (2-7)$$

которая, как следует из предыдущего обсуждения, приводит к сдвигу изображения. Преобразование векторов A и B с помощью матрицы $[T]$ дает новое положение векторов A^* и B^*

$$[A][T] = [0 \ 1] \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} = [3 \ 1] = [A^*] \quad (2-8)$$

и

$$[B][T] = [2 \ 3] \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} = [11 \ 7] = [B^*]. \quad (2-9)$$

Таким образом, результирующие координаты для точки A^* — это $x^* = 3$ и $y^* = 1$. Аналогично, B^* — новая точка с координатами $x^* = 11$, $y^* = 7$. В более компактном виде отрезок AB может быть представлен матрицей размером (2×2) :

$$L = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}.$$

Умножим эту матрицу на $[T]$:

$$[L][T] = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 11 & 7 \end{bmatrix} = [L^*], \quad (2-10)$$

где компоненты $[L^*]$ представляют собой преобразование координат векторов $[A^*]$ и $[B^*]$. Результат преобразования A в A^* и B в B^* показан на рис. 2-2, где x и y — это исходные оси координат, а x^* и y^* — преобразованные оси. Из рисунка видно, что преобразование сдвига $[T]$ увеличивает длину отрезка и изменяет его направление.

2-6 ПРЕОБРАЗОВАНИЕ СРЕДНЕЙ ТОЧКИ

На рис. 2-2 видно, что (2×2) -матрица преобразует прямую линию $y = x + 1$, проходящую между точками A и B в другую прямую $y = \frac{3}{4}x - \frac{5}{4}$, которая проходит между точками A^* и B^* . Фактически с помощью (2×2) -матрицы осуществляется преобразование любой прямой в другую прямую. Все точки преобразованной линии непосредственно соответствуют всем точкам исходной линии. Это достаточно очевидно для конечных точек линии. Рассмотрим теперь преобразование средней точки прямой линии AB . Допустим, что

$$[A] = [x_1 \ y_1], \quad [B] = [x_2 \ y_2] \quad \text{и} \quad [T] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Преобразуем одновременно две крайние точки:

$$\begin{aligned} \begin{bmatrix} A \\ B \end{bmatrix} [T] &= \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \\ &= \begin{bmatrix} ax_1 + cy_1 & bx_1 + dy_1 \\ ax_2 + cy_2 & bx_2 + dy_2 \end{bmatrix} = \begin{bmatrix} A^* \\ B^* \end{bmatrix}, \end{aligned} \quad (2-11)$$

Итак, конечные точки преобразованной линии A^*B^* имеют следующие координаты

$$\begin{aligned} [A^*] &= [ax_1 + cy_1 \ bx_1 + dy_1] = [x_1^* \ y_1^*], \\ [B^*] &= [ax_2 + cy_2 \ bx_2 + dy_2] = [x_2^* \ y_2^*]. \end{aligned} \quad (2-12)$$

Средняя точка отрезка A^*B^* выражается через преобразованные конечные точки

$$\begin{aligned} [x_m^* \ y_m^*] &= \left[\frac{x_1^* + x_2^*}{2} \ \frac{y_1^* + y_2^*}{2} \right] = \\ &= \left[\frac{(ax_1 + cy_1) + (ax_2 + cy_2)}{2} \ \frac{(bx_1 + dy_1) + (bx_2 + dy_2)}{2} \right] = \\ &= \left[a \frac{(x_1 + x_2)}{2} + c \frac{(y_1 + y_2)}{2} \ b \frac{(x_1 + x_2)}{2} + d \frac{(y_1 + y_2)}{2} \right]. \end{aligned} \quad (2-13)$$

Возвращаясь к исходной линии AB , можно определить среднюю точку следующим образом:

$$[x_m \ y_m] = [(x_1 + x_2)/2 \ (y_1 + y_2)/2]. \quad (2-14)$$

Применив матрицу преобразования $[T]$ к средней точке линии AB , получаем:

$$\begin{bmatrix} x_m & y_m \end{bmatrix} [T] = \begin{bmatrix} \frac{x_1 + x_2}{2} & \frac{y_1 + y_2}{2} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a \frac{(x_1 + x_2)}{2} + c \frac{(y_1 + y_2)}{2} & b \frac{(x_1 + x_2)}{2} + d \frac{(y_1 + y_2)}{2} \end{bmatrix}. \quad (2-15)$$

Из сравнения выражений (2-13) и (2-15) видно, что они одинаковы, и поэтому средняя точка линии AB преобразуется в среднюю точку линии A^*B^* . Такой метод можно применить и к любым другим отрезкам разделенной линии. Таким образом, при преобразовании путем умножения на матрицу гарантируется соответствие всех точек линии AB и A^*B^* .

Пример 2-1 Средняя точка прямой

Рассмотрим отрезок AB из рис. 2-2. Положение векторов конечных точек такое: $[A] = [0 \ 1]$, $[B] = [2 \ 3]$. Преобразование $[T] = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$ осуществляет перемещение вектора на линию A^*B^* :

$$\begin{bmatrix} A \\ B \end{bmatrix} [T] = \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 11 & 7 \end{bmatrix} = \begin{bmatrix} A^* \\ B^* \end{bmatrix}.$$

Средняя точка A^*B^* будет иметь координаты

$$\begin{bmatrix} x_m^* & y_m^* \end{bmatrix} = \left[\frac{3+11}{2} \quad \frac{1+7}{2} \right] = [7 \ 4].$$

Координаты средней точки линии AB равны

$$\begin{bmatrix} x_m & y_m \end{bmatrix} = \left[\frac{0+2}{2} \quad \frac{1+3}{2} \right] = [1 \ 2].$$

Преобразуем среднюю точку и получим

$$\begin{bmatrix} x_m & y_m \end{bmatrix} [T] = [1 \ 2] \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} = [7 \ 4] = \begin{bmatrix} x_m^* & y_m^* \end{bmatrix},$$

что полностью эквивалентно предыдущему результату.

Применением этих результатов в машинной графике любая прямая может быть преобразована в любую другую прямую путем простого преобразования ее конечных точек и восстановления линии между ними.

2-7 ПРЕОБРАЗОВАНИЕ ПАРАЛЛЕЛЬНЫХ ЛИНИЙ

Результатом преобразования двух параллельных линий с помощью (2×2) -матрицы снова будут две параллельные линии. Это можно увидеть, рассмотрев линию между точками $[A] = [x_1 \ y_1]$, $[B] = [x_2 \ y_2]$ и параллельную ей линию,

2-8 Преобразование пересекающихся прямых

проходящую между точками E и F . Покажем, что для этих линий любое преобразование сохраняет параллельность. Так как AB , EF и A^*B^* и E^*F^* параллельны, то угол наклона линий AB и EF определяется следующим образом:

$$m = \frac{y_2 - y_1}{x_2 - x_1}. \quad (2-16)$$

Преобразуем конечные точки AB , воспользовавшись матрицей общего преобразования размером (2×2) :

$$\begin{bmatrix} A \\ B \end{bmatrix} [T] = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ax_1 + cy_1 & bx_1 + dy_1 \\ ax_2 + cy_2 & bx_2 + dy_2 \end{bmatrix} = \begin{bmatrix} x_1^* & y_1^* \\ x_2^* & y_2^* \end{bmatrix} = \begin{bmatrix} A^* \\ B^* \end{bmatrix}. \quad (2-17)$$

Наклон прямой A^*B^* определяется следующим образом:

$$m^* = \frac{(bx_2 + dy_2) - (bx_1 + dy_1)}{(ax_2 + cy_2) - (ax_1 + cy_1)} = \frac{b(x_2 - x_1) + d(y_2 - y_1)}{a(x_2 - x_1) + c(y_2 - y_1)}$$

или

$$m^* = \frac{b + d \frac{(y_2 - y_1)}{(x_2 - x_1)}}{a + c \frac{(y_2 - y_1)}{(x_2 - x_1)}} = \frac{b + dm}{a + cm}. \quad (2-18)$$

Так как наклон m^* не зависит от x_1 , x_2 , y_1 , y_2 а m , a , b , c и d одинаковы для EF и AB , то m^* одинаково для E^*F^* и A^*B^* . Таким образом, параллельные линии сохраняют параллельность и после преобразования. Это означает, что при преобразовании (2×2) параллелограмм преобразуется в другой параллелограмм. Эти тривиальные выводы демонстрируют большие возможности использования матрицы преобразования для создания графических эффектов.

2-8 ПРЕОБРАЗОВАНИЕ ПЕРЕСЕКАЮЩИХСЯ ПРЯМЫХ

Результатом преобразования с помощью (2×2) -матрицы пары пересекающихся прямых линий также будет пара пересекающихся линий. Проиллюстрируем этот факт на примере двух прямых, изображенных на рис. 2-3 штриховой линией и заданных уравнениями

$$\begin{aligned} y &= m_1 x + b_1, \\ y &= m_2 x + b_2. \end{aligned}$$

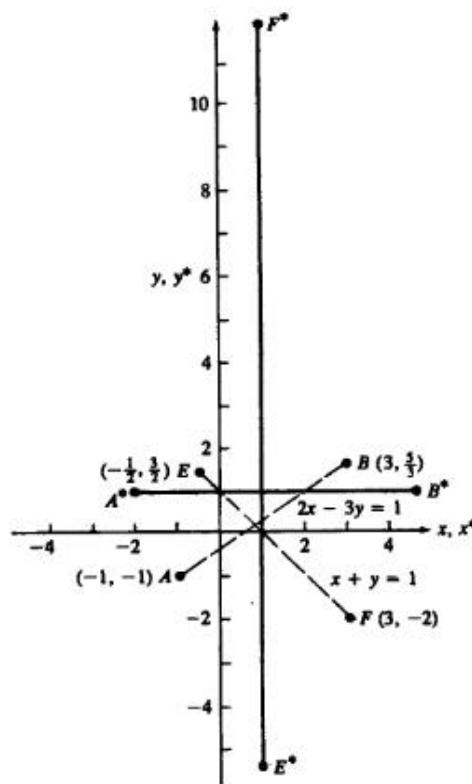


Рис. 2-3 Преобразование пересекающихся прямых.

В матричном представлении эти уравнения будут иметь вид:

$$[x \ y] \begin{bmatrix} -m_1 & -m_2 \\ 1 & 1 \end{bmatrix} = [b_1 \ b_2]$$

или

$$[X][M] = [B]. \quad (2-19)$$

Если существует решение этой системы уравнений, то линии пересекаются, в противном случае они параллельны. Решение можно найти путем инверсии матрицы. В частности,

$$[X_i] = [x_i \ y_i] = [B][M]^{-1}. \quad (2-20)$$

Матрица, обратная $[M]$, имеет следующий вид:

$$[M]^{-1} = \begin{bmatrix} 1 & m_2 \\ \frac{m_2 - m_1}{m_2 - m_1} & \frac{m_2 - m_1}{m_2 - m_1} \end{bmatrix}, \quad (2-21)$$

так как $[M][M]^{-1} = [I]$, где $[I]$ — единичная матрица. Поэтому координаты точки пересечения двух линий можно найти следующим образом:

$$\begin{aligned} [X_i] &= [x_i \ y_i] = [b_1 \ b_2] \begin{bmatrix} 1 & m_2 \\ \frac{m_2 - m_1}{m_2 - m_1} & \frac{m_2 - m_1}{m_2 - m_1} \end{bmatrix}, \\ [X_i] &= [x_i \ y_i] = \left[\frac{b_1 - b_2}{m_2 - m_1} \quad \frac{b_1 m_2 - b_2 m_1}{m_2 - m_1} \right], \end{aligned} \quad (2-22)$$

Если обе линии преобразовать с помощью (2×2) -матрицы общего преобразования вида

$$[T] = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

то их уравнения будут иметь вид

$$\begin{aligned} y^* &= m_1^* x^* + b_1^*, \\ y^* &= m_2^* x^* + b_2^*. \end{aligned}$$

Соответственно можно показать, что

$$m_i^* = \frac{b + dm_i}{a + cm_i} \quad (2-23)$$

и

$$b_i^* = b_i(d - cm_i) = b_i \frac{ad - bc}{a + cm_i}, \quad \text{где } i = 1, 2. \quad (2-24)$$

Точка пересечения линий после преобразования отыскивается таким же образом, как и в случае исходных линий:

$$\begin{aligned} [X_i^*] &= [x_i^* \ y_i^*] = \\ &= \left[\frac{b_1^* - b_2^*}{m_2^* - m_1^*} \quad \frac{b_1^* m_2^* - b_2^* m_1^*}{m_2^* - m_1^*} \right]. \end{aligned}$$

Воспользовавшись выражениями (2-23) и (2-24), получим

$$\begin{aligned} [X_i^*] &= [x_i^* \ y_i^*] = \\ &= \left[\frac{a(b_1 - b_2) + c(b_1 m_2 - b_2 m_1)}{m_2 - m_1} \quad \frac{b(b_1 - b_2) + d(b_1 m_2 - b_2 m_1)}{m_2 - m_1} \right]. \end{aligned} \quad (2-25)$$

Возвращаясь теперь к точке пересечения $[x_i \ y_i]$ исходных линий и применяя уже полученную матрицу преобразования, имеем

$$\begin{aligned} [x_i^* \ y_i^*] &= [x_i \ y_i] [T] = \\ &= \left[\frac{b_1 - b_2}{m_2 - m_1} \quad \frac{b_1 m_2 - b_2 m_1}{m_2 - m_1} \right] \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \\ &= \left[\frac{a(b_1 - b_2) + c(b_1 m_2 - b_2 m_1)}{m_2 - m_1} \quad \frac{b(b_1 - b_2) + d(b_1 m_2 - b_2 m_1)}{m_2 - m_1} \right]. \end{aligned} \quad (2-26)$$

Сравнение уравнений (2-25) и (2-26) показывает, что они одинаковы. Итак, точка пересечения преобразуется точно в другую точку пересечения.

Пример 2-2 Пересекающиеся прямые

Рассмотрим две штриховые линии AB и EF на рис. 2-3, конечные точки которых имеют координаты

$$[A] = [-1 \ -1], \quad [B] = [3 \ 5/3]$$

и

$$[E] = [-1/2 \ 3/2], \quad [F] = [3 \ -2].$$

Уравнение прямой AB имеет вид $-(2/3)x + y = -1/3$, а прямая EF задается уравнением $x + y = 1$. В матричном виде пучок прямых представляется в виде

$$[x \ y] \begin{bmatrix} -2/3 & 1 \\ 1 & 1 \end{bmatrix} = [-1/3 \ 1].$$

Используя матрицу обратного преобразования (2-21), получим точку пересечения этих прямых

$$[x_i \ y_i] = [-1/3 \ 1] \begin{bmatrix} -3/5 & -3/5 \\ 3/5 & 2/5 \end{bmatrix} = [4/5 \ 1/5].$$

Теперь преобразуем эти линии с помощью матрицы

$$[T] = \begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix}.$$

Результирующие прямые A^*B^* и E^*F^* показаны на рис. 2-3. В матричном виде уравнения преобразованных линий имеют вид

$$[x^* \ y^*] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [1 \ 1]$$

с точкой пересечения $[x_i^* \ y_i^*] = [1 \ 1]$.

Преобразуя точку пересечения исходных линий, получим

$$[x_i^* \ y_i^*] = [x_i \ y_i] [T] = [4/5 \ 1/5] \begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix} = [1 \ 1],$$

что тождественно точке пересечения преобразованных линий.

Из рис. 2-3 и примера 2-2 видно, что исходные штриховые прямые AB и EF не перпендикулярны друг другу. Однако преобразованные прямые A^*B^* и E^*F^* , показанные сплошной линией, являются перпендикулярными. Таким образом, преобразование $[T]$ переводит две пересекающиеся неперпендикулярные прямые в две пересекающиеся перпендикулярные. Смысл обратного преобразования $[T]^{-1}$

состоит в переводе двух пересекающихся перпендикулярных прямых в две пересекающиеся, но не перпендикулярные, что может привести к неприятным геометрическим последствиям. Значительный интерес представляет вопрос, при каком условии перпендикулярные прямые преобразуются в перпендикулярные. Мы вернемся к этому вопросу в разд. 2-14, где разберем его подробнее.

Дополнительное изучение рис. 2-3 и примера 2-2 показывает, что преобразование $[T]$ включает в себя поворот, отражение и масштабирование. Рассмотрим каждое из этих преобразований отдельно.

2-9 ПОВОРОТ

Рассмотрим треугольник ABC (рис. 2-4) и с помощью следующего преобразования повернем его на 90° против часовой стрелки относительно начала координат

$$[T] = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Если использовать матрицу размером (3×2) , состоящую из координат x и y вершин треугольника, то можно записать

$$\begin{bmatrix} 3 & -1 \\ 4 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -1 & 4 \\ -1 & 2 \end{bmatrix},$$

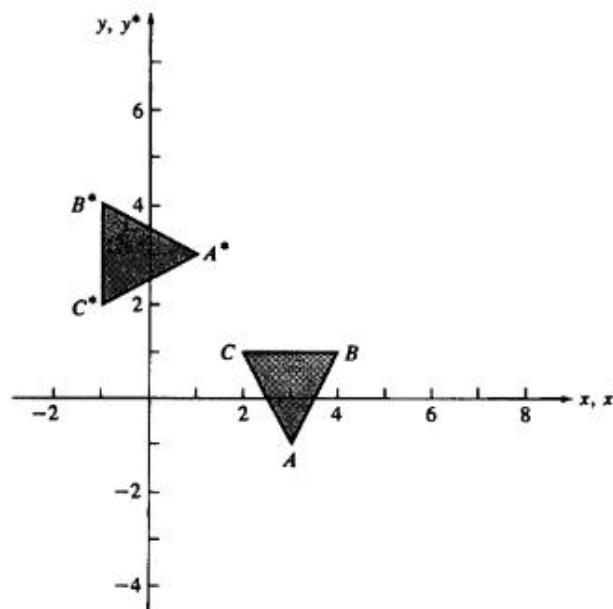


Рис. 2-4 Поворот.

что является координатами результирующего треугольника $A^*B^*C^*$. Поворот на 180° относительно начала координат достигается путем следующего преобразования

$$[T] = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

а на 270° относительно начала координат — преобразованием

$$[T] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Разумеется, что матрица тождественного преобразования

$$[T] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

соответствует повороту вокруг начала координат на 0° или 360° . Обратим внимание, что в этих примерах не встречаются ни масштабирование, ни отражение.

В этих примерах осуществляется преобразование в специальных случаях поворота вокруг начала координат на углы $0^\circ, 90^\circ, 180^\circ$ и 270° . Как осуществить поворот вокруг точки начала координат на произвольный угол θ ? Для ответа на этот вопрос рассмотрим вектор положения от начала координат до точки P (рис. 2-5). Обозначим r — длину вектора, а ϕ — угол между вектором и осью x . Вектор положения поворачивается вокруг начала координат на угол θ и попадает в точку P^* . Записав векторы положений для P и P^* , получаем:

$$P = [x \ y] = [r \cos \phi \ r \sin \phi]$$

и

$$P^* = [x^* \ y^*] = [r \cos(\phi + \theta) \ r \sin(\phi + \theta)].$$

Используя формулу для \cos суммы углов, перепишем выражение для P^* следующим образом

$$P^* = [x^* \ y^*] = [r(\cos \phi \cos \theta - \sin \phi \sin \theta) \ r(\cos \phi \sin \theta + \sin \phi \cos \theta)].$$

Используя определения x и y , можно переписать P^* как

$$P^* = [x^* \ y^*] = [x \cos \theta - y \sin \theta \ x \sin \theta + y \cos \theta].$$

Таким образом, преобразованная точка имеет координаты

$$x^* = x \cos \theta - y \sin \theta, \quad (2-27a)$$

$$y^* = x \sin \theta + y \cos \theta. \quad (2-27b)$$

или в матричном виде

$$\begin{aligned} [X^*] &= [X][T] = [x^* \ y^*] = \\ &= [x \ y] \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. \end{aligned} \quad (2-28)$$

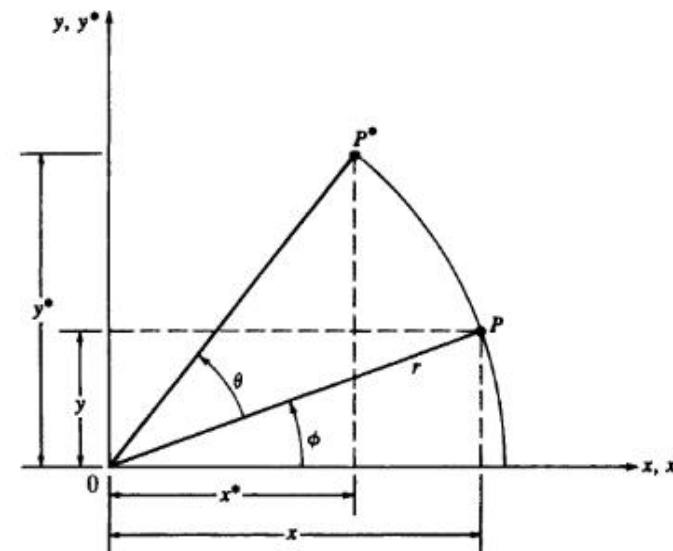


Рис. 2-5 Поворот координатного вектора.

Итак, преобразование поворота вокруг точки начала координат на произвольный угол θ задается матрицей

$$[T] = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. \quad (2-29)$$

Повороты являются положительными, если они осуществляются против часовой стрелки относительно точки вращения (рис. 2-5).

Определитель общей матрицы поворота имеет следующий вид:

$$\det[T] = \cos^2 \theta + \sin^2 \theta = 1. \quad (2-30)$$

В общем случае преобразования по матрице с детерминантом, равным 1, приводят к полному повороту.

Предположим теперь, что требуется возвратить точку P^* обратно в P , т. е. выполнить обратное преобразование. Очевидно, что требуемый угол поворота равен $-\theta$. Из формулы (2-29) возьмем матрицу для выполнения необходимого преобразования

$$[T]^{-1} = \begin{bmatrix} \cos(-\theta) & \sin(-\theta) \\ -\sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad (2-31)$$

так как $\cos(-\theta) = \cos \theta$ и $\sin(-\theta) = -\sin \theta$. Выражение $[T]^{-1}$ является формальной записью обратной матрицы $[T]$. Можно показать, что матрица $[T]^{-1}$ является

обратной к $[T]$, если вспомнить, что результат умножения матрицы на обратную дает единичную матрицу. В нашем случае:

$$\begin{aligned}[T][T]^{-1} &= \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \\ &= \begin{bmatrix} \cos^2 \theta + \sin^2 \theta & -\cos \theta \sin \theta + \cos \theta \sin \theta \\ -\cos \theta \sin \theta + \cos \theta \sin \theta & \cos^2 \theta + \sin^2 \theta \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [I],\end{aligned}$$

где $[I]$ — единичная матрица.

Анализ выражений (2-29) и (2-31) приводит к другому интересному и полезному результату. Вспомним, что транспонирование матрицы определяется заменой ее строк столбцами. Обозначим транспонированную матрицу $[T]$ как $[T]^T$. Сравнивая ее с $[T]^{-1}$, видим, что

$$[T]^T = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = [T]^{-1}. \quad (2-32)$$

Обратная матрица вращения является транспонированной. Поскольку формально определитель обратной матрицы вычисляется гораздо сложнее, чем определитель транспонированной, то выражение (2-32) является достаточно важным и полезным результатом. В общем случае обратной для любой матрицы преобразования полного поворота, т. е. матрицы с определителем, равным +1, является ее транспонированная матрица (такие матрицы называют ортогональными).

2-10 ОТРАЖЕНИЕ

В то время как полный поворот на плоскости xy обычно осуществляется в двумерном пространстве относительно нормали к плоскости, отражение представляет собой тот же поворот на угол 180° в трехмерном пространстве и обратно на плоскость относительно оси, лежащей в плоскости xy . На рис. 2-6 приведены примеры двух отражений на плоскости треугольника DEF . Отражение относительно прямой $y = 0$ (ось x) получено с использованием матрицы

$$[T] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2-33)$$

В этом случае новые вершины треугольника $D^*E^*F^*$ будут определяться преобразованием

$$\begin{bmatrix} 8 & 1 \\ 7 & 3 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 8 & -1 \\ 7 & -3 \\ 6 & -2 \end{bmatrix}.$$

Подобным образом отражение относительно оси y при $x = 0$ будет иметь вид

$$[T] = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (2-34)$$

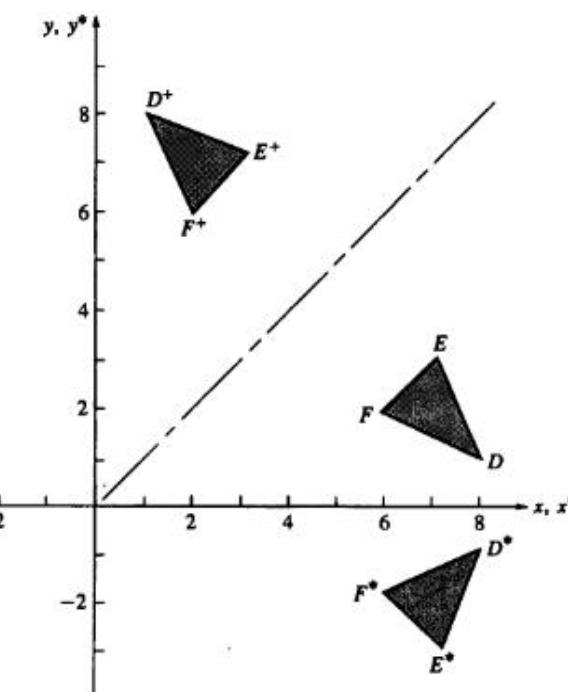


Рис. 2-6 Отражение.

Отражение относительно прямой $y = x$ осуществляется с помощью матрицы

$$[T] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (2-35)$$

Выполнив преобразования, получим координаты вершин треугольника $D^*E^*F^*$

$$\begin{bmatrix} 8 & 1 \\ 7 & 3 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 8 \\ 3 & 7 \\ 2 & 6 \end{bmatrix}.$$

Аналогичным образом отражение относительно оси x будет иметь вид

$$[T] = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}. \quad (2-36)$$

У каждой из этих матриц определитель равен -1 . В общем случае, если определитель матрицы преобразования равен -1 , то преобразование дает полное отражение.

Если оба полных отражения осуществляются последовательно относительно прямых, проходящих через начало координат, то результатом будет полный поворот относительно начала координат. Это можно увидеть, обратившись к следующему примеру.

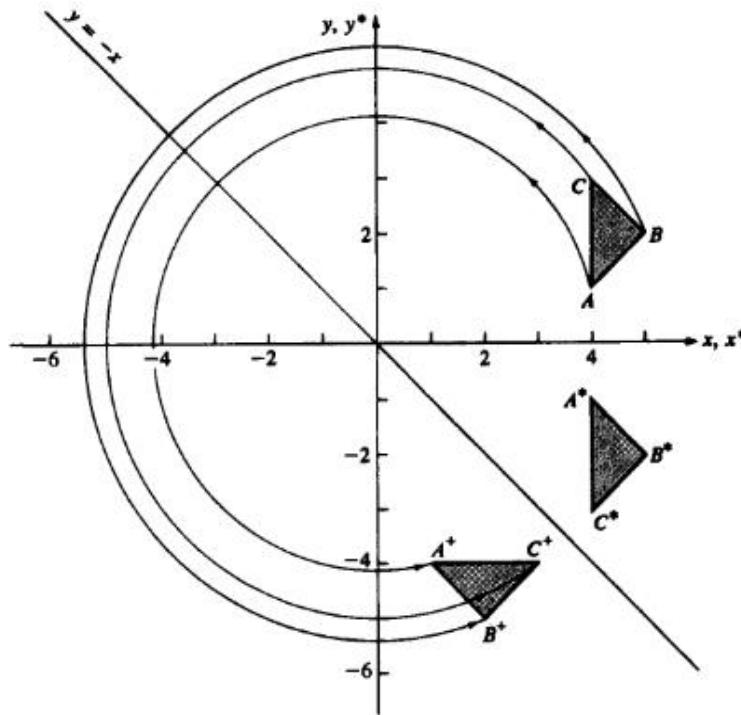


Рис. 2-7 Выполнение отражения путем поворота.

Пример 2-3 Отражение и вращение

Рассмотрим треугольник ABC , показанный на рис. 2-7. Первоначально отобразим его относительно оси x (уравнение 2-33), а затем относительно прямой $y = -x$ (см. выражение (2-36)). Результатом первого отображения будет

$$[X^*] = [X][T_1] = \begin{bmatrix} 4 & 1 \\ 5 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 5 & -2 \\ 4 & -3 \end{bmatrix}.$$

Результатом второго будет

$$[X^{**}] = [X^*][T_2] = \begin{bmatrix} 4 & -1 \\ 5 & -2 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ 2 & -5 \\ 3 & -4 \end{bmatrix}.$$

Повернем треугольник относительно начала координат на угол $\theta = 270^\circ$ (см. (2-29)) и получим аналогичный результат

$$[X^{**}] = [X][T_3] = \begin{bmatrix} 4 & 1 \\ 5 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ 2 & -5 \\ 3 & -4 \end{bmatrix}.$$

Отметим, что матрицы отражения из (2-33) и (2-36) ортогональны, т. е. транспонированная матрица одновременно является обратной. Например,

$$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}^T = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}^{-1}.$$

2-11 МАСШТАБИРОВАНИЕ

Из наших рассуждений относительно преобразования точек следует, что величина масштабирования определяется значением элементов исходной диагональной матрицы. Если матрица

$$[T] = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

используется в качестве оператора воздействия на вершины треугольника, то имеет место «двукратное» расширение или равномерное масштабирование относительно точки начала координат. Если значения элементов не равны, то треугольник искажается, что проиллюстрировано на рис. 2-8. Треугольник ABC , преобразованный с помощью матрицы

$$[T] = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

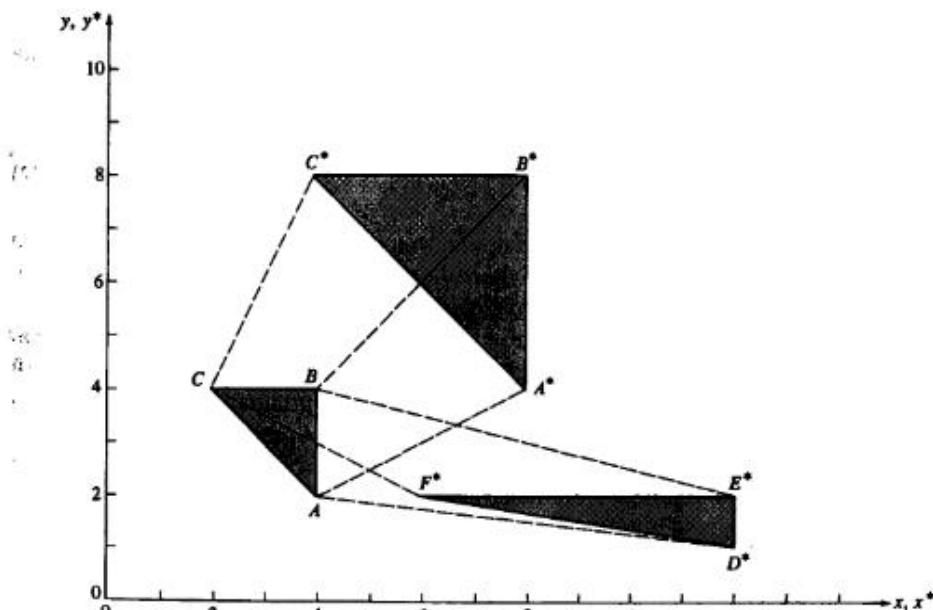


Рис. 2-8 Пропорциональное и непропорциональное масштабирование (искажение).

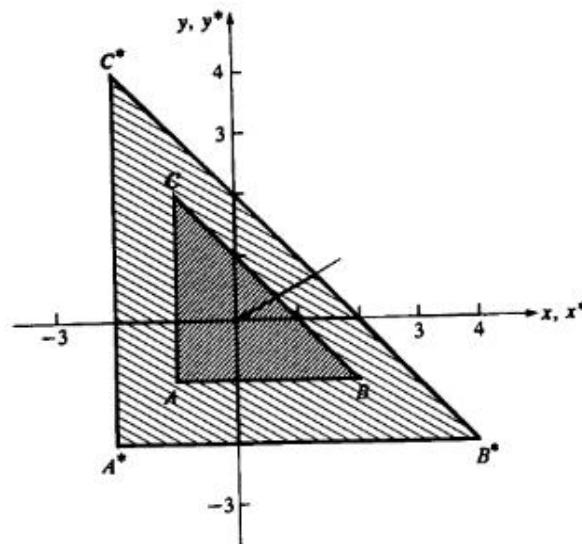


Рис. 2-9 Пропорциональное масштабирование без явного перемещения.

переходит в пропорционально увеличенный треугольник $A^*B^*C^*$. Тот же треугольник, но преобразованный с помощью матрицы

$$\begin{bmatrix} 1/2 & 0 \\ 0 & 3 \end{bmatrix}$$

переходит в треугольник $D^*E^*F^*$, имеющий искажение, вызванное разными коэффициентами масштабирования.

В общем случае при матрице

$$[T] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad (2-37)$$

в которой $a = d, b = c = 0$, выполняется пропорциональное масштабирование; если $a \neq d, b = c = 0$, то масштабирование будет проведено непропорционально. В первом случае для $a = d > 1$ происходит расширение, т. е. увеличение изображения. Если $a = d < 1$, то происходит равномерное сжатие, т. е. фигура уменьшается. Непропорциональное расширение и сжатие возникают в зависимости от значений a и d , которые могут быть меньше либо больше, чем 1, независимо друг от друга.

Из рис. 2-8 видно также, что на первый взгляд преобразование треугольника является перемещением. Это объясняется тем, что относительно начала координат масштабируются координатные векторы, а не точки.

Для того чтобы лучше понять этот факт, рассмотрим преобразования ABC в $D^*E^*F^*$ более внимательно. В частности,

$$[X^*] = [X][T] = \begin{bmatrix} 4 & 2 \\ 4 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 2 & 12 \\ 1 & 12 \end{bmatrix}.$$

Заметим, что каждая из компонент x координатных векторов треугольника DEF умножалась на масштабный коэффициент 3, а компоненты y — на 2.

Для того чтобы получить чистое масштабирование без эффекта перемещения, центр фигуры надо поместить в начало координат. Это видно из рис. 2-9, на котором треугольник ABC увеличивается в два раза при масштабировании относительно его центра с координатами, равными $1/3$ основания и $1/3$ высоты. Конкретная матрица преобразования имеет вид

$$[X^*] = [X][T] = \begin{bmatrix} -1 & -1 \\ 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 4 & -2 \\ -2 & 4 \end{bmatrix}.$$

2-12 КОМБИНИРОВАННЫЕ ПРЕОБРАЗОВАНИЯ

Возможности матричного метода достаточно ясно описаны в предыдущих разделах книги. С помощью матричных операций над координатными векторами, определяющими вершины фигур, можно управлять формой и положением поверхности. Однако для получения желаемой ориентации может потребоваться более одного преобразования. Так как операция умножения матриц не коммутативна, то важен порядок выполнения преобразования.

Для иллюстрации эффекта некоммутативности операции умножения матриц рассмотрим преобразования поворота и отражения координатного вектора $[x \ y]$. Если вслед за поворотом на 90° (посредством $[T_1]$) производится отражение относительно прямой $y = -x$ (посредством $[T_2]$), то эти два последовательных преобразования дают

$$[X'] = [X][T_1] = [x \ y] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = [-y \ x]$$

и затем

$$[X^*] = [X'][T_2] = [-y \ -x] \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = [-x \ y].$$

С другой стороны, если отражение следует за поворотом, то получатся следующие результаты:

$$[X'] = [X][T_2] = [x \ y] \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = [-y \ -x]$$

и

$$[X^*] = [X'][T_1] = [-y \ -x] \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = [x \ -y].$$

Оба результата различны, что подтверждает важность порядка применения матричных преобразований.

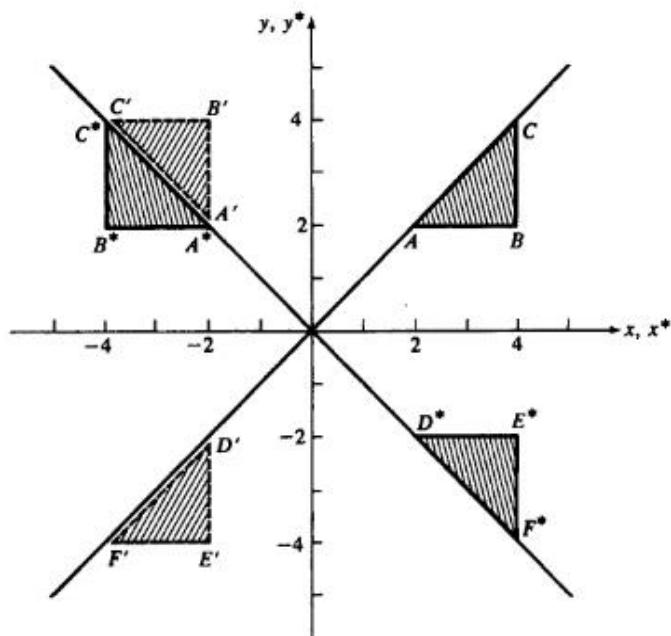


Рис. 2-10 Комбинированное преобразование на плоскости.

Другое принципиальное обстоятельство иллюстрируется этими результатами и приведенным ниже примером. Ранее отдельные матричные преобразования применялись последовательно к результатам предыдущих преобразований, например

$$[x \ y] [T_1] \rightarrow [x' \ y']$$

и

$$[x' \ y'] [T_2] \rightarrow [x^* \ y^*].$$

В приведенном ниже примере отдельные преобразования предварительно комбинируются или конкатенируются, а затем полученная матрица применяется к исходному вектору, т. е. $[T_1][T_2] \rightarrow [T_3]$ и $[x \ y] [T_3] \rightarrow [x^* \ y^*]$.

Пример 2-4 Комбинированные преобразования на плоскости

Рассмотрим треугольник ABC на рис. 2-10. Выполним над ним два преобразования: поворот на $+90^\circ$ вокруг точки начала координат

$$[T_1] = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

и отражение относительно линии $y = -x$

$$[T_2] = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Результатом воздействия комбинированного преобразования $[T_3] = [T_1][T_2]$ на треугольник ABC является

$$[X^*] = [X][T_1][T_2] = [X][T_3]$$

или

$$\begin{bmatrix} 2 & 2 \\ 4 & 2 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 4 & 2 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ -4 & 2 \\ -4 & 4 \end{bmatrix}.$$

Получившийся треугольник $A^*B^*C^*$ является конечным результатом данного преобразования, а треугольник $A'B'C'$ — промежуточным результатом (рис. 2-10).

Проведем преобразование в обратном порядке

$$[X^*] = [X][T_2][T_1] = [X][T_4]$$

или

$$\begin{bmatrix} 2 & 2 \\ 4 & 2 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 4 & 2 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 4 & -2 \\ 4 & -4 \end{bmatrix}.$$

Конечным результатом будет треугольник $D'E^*F^*$, а промежуточным $D'E'F'$ (рис. 2-10). Оба результата различны, тем самым снова подтверждается важность порядка применения преобразований. Отметим также, что для определителей справедливы равенства $\det[T_3] = -1$, и $\det[T_4] = -1$ и поэтому оба результата могут быть получены с помощью единственного отражения. Треугольник $A^*B^*C^*$ можно получить из ABC путем отражения относительно оси Y (матрица $[T_3]$ и уравнение (2-34), $D'E^*F^*$ получается из ABC при отражении относительно оси X (матрица $[T_4]$ и уравнение (2-33)).

2-13 ПРЕОБРАЗОВАНИЕ ЕДИНИЧНОГО КВАДРАТА

До сих пор мы рассматривали поведение точек и линий для определения результатов простых матричных преобразований. Однако можно корректно рассматривать применение матрицы к любой точке плоскости. Как было показано ранее, единственная точка, остающаяся инвариантной при воздействии матричных преобразований, — это точка начала координат. Все другие точки плоскости подвержены преобразованию, которое можно представить как растяжение исходной плоскости, системы координат и перевод в новую форму. Формально принято считать, что преобразование вызывает переход от одного координатного пространства к другому.

Рассмотрим координатную сетку, состоящую из единичных квадратов на координатной плоскости xy (рис. 2-11). Четыре координатных вектора вершин единичного квадрата, проходящие под одним углом к началу координат, имеют следующий вид:

$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$	начало координат — A
$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$	единичная точка B на оси x
$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	внешний угол C
$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	единичная точка D на оси y

Такой единичный квадрат изображен на рис. 2-11a. Применяя к нему (2×2) -матрицу общего преобразования, получаем

$$\begin{matrix} A & \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \\ B & \begin{bmatrix} a & b \\ c & d \end{bmatrix} \end{matrix} = \begin{matrix} A^* \\ B^* \\ C^* \\ D^* \end{matrix} = \begin{bmatrix} 0 & 0 \\ a & b \\ a+c & b+d \\ c & d \end{bmatrix} \quad (2-38)$$

Результаты этого преобразования показаны на рис. 2-11b. Из выражения (2-38) следует, что начало координат не подвергается преобразованию, т. е. $[A] = [A^*] = [0 \ 0]$. Далее отметим, что координаты B^* равны первой строке матрицы преобразования, а координаты D^* — второй. Таким образом, матрица преобразования является определенной, если определены координаты B^* и D^* (преобразование единичных векторов $[1 \ 0]$, $[0 \ 1]$). Поскольку стороны единичного квадрата первоначально параллельны и ранее было показано, что параллельные линии преобразуются снова в параллельные, то результатирующая фигура является параллелограммом.

Влияние элементов a , b , c и d матрицы 2×2 может быть установлено отдельно. Элементы b и c , как видно из рис. 2-11b, вызывают сдвиг (см. разд. 2-4) исходного квадрата в направлениях y и x соответственно. Как отмечалось ранее, элементы a и d играют роль масштабных множителей. Таким образом, 2×2 -матрица задает комбинацию сдвига и масштабирования.

Несложно определить также площадь параллелограмма $A^*B^*C^*D^*$ из рис. 2-11b, которую можно вычислить следующим образом:

$$A_p = (a+c)(b+d) - \frac{1}{2}(ab) - \frac{1}{2}(cd) - \frac{c}{2}(b+b+d) - \frac{b}{2}(c+a+c).$$

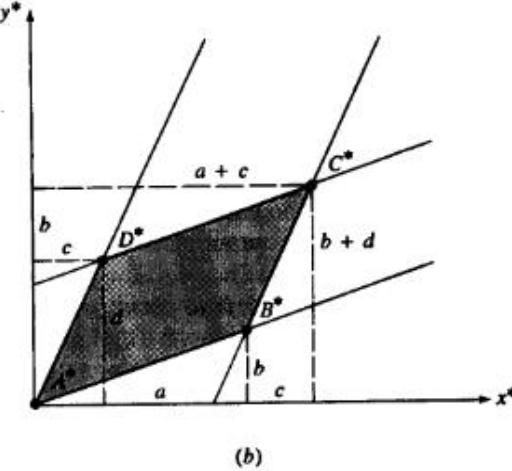
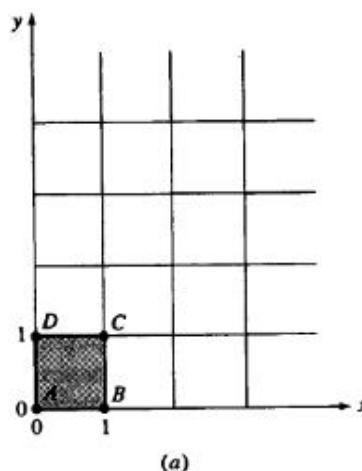


Рис. 2-11 Общее преобразование единичного квадрата: a) до преобразования; b) после преобразования.

В результате получаем

$$A_p = ad - bc = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \quad (2-39)$$

Можно показать, что площадь любого параллелограмма A_p , образованного путем преобразования квадрата, есть функция от определителя матрицы преобразования и связана с площадью исходного квадрата A_s простым отношением

$$A_p = A_s(ad - bc) = A_s \det [T]. \quad (2-40)$$

Фактически, так как площадь всей фигуры равна сумме площадей единичных квадратов, то площадь любой преобразованной фигуры A_t зависит от площади исходной фигуры A_i

$$A_t = A_i(ad - bc). \quad (2-41)$$

Это полезный способ определения площадей произвольных фигур.

Пример 2-5 Масштабирование области

Треугольник ABC с координатными векторами $[1 \ 0]$, $[0 \ 1]$ и $[-1 \ 0]$ преобразуется матрицей

$$[T] = \begin{bmatrix} 3 & 2 \\ -1 & 2 \end{bmatrix},$$

образуя новый треугольник $A^*B^*C^*$ (рис. 2-12).

Площадь треугольника ABC равна

$$A_i = \frac{1}{2} (\text{основание}) (\text{высота}) = \frac{1}{2}(2)(1) = 1.$$

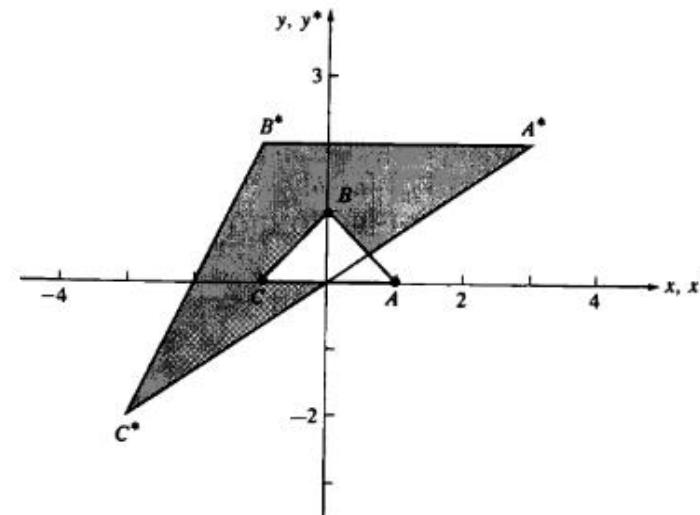


Рис. 2-12 Масштабирование области.

Воспользуемся уравнением (2-41), тогда площадь преобразованного треугольника $A^*B^*C^*$ будет равна

$$A_t = A_i(ad - bc) = 1(6 + 2) = 8.$$

Векторы преобразованного треугольника $A^*B^*C^*$ теперь равны

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -1 & 2 \\ -3 & -2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ -1 & 2 \\ -3 & -2 \end{bmatrix}.$$

Вычислим площадь, образованную результирующими вершинами:

$$A_t = \frac{1}{2}(\text{основание})(\text{высота}) = \frac{1}{2}(4)(4) = 8.$$

Это совпадает с полученным ранее результатом.

2-14 ПРЕОБРАЗОВАНИЯ ЖЕСТКИХ КОНСТРУКЦИЙ

Пришло время вернуться к поставленному в разд. 2-8 вопросу: когда перпендикулярные прямые преобразуются в перпендикулярные прямые? Рассмотрим сначала более общий вопрос: в каких случаях угол между пересекающимися прямыми сохраняется?

Напомним, что скалярное произведение двух векторов равно

$$\bar{V}_1 \cdot \bar{V}_2 = V_{1x}V_{2x} + V_{1y}V_{2y} = |\bar{V}_1| |\bar{V}_2| \cos \theta, \quad (2-42)$$

а векторное произведение двух векторов, принадлежащих плоскости xy , определяется как

$$\bar{V}_1 \times \bar{V}_2 = (V_{1x}V_{2y} - V_{2x}V_{1y})\bar{k} = |\bar{V}_1| |\bar{V}_2| \bar{k} \sin \theta, \quad (2-43)$$

где индексы x и y относятся к компонентам x и y вектора, θ — острый угол между векторами, а \bar{k} — единичный вектор, перпендикулярный к плоскости xy .

Проведем преобразование \bar{V}_1 и \bar{V}_2 , используя (2×2) -матрицу общего преобразования

$$\begin{bmatrix} \bar{V}_1 \\ \bar{V}_2 \end{bmatrix} [T] = \begin{bmatrix} V_{1x} & V_{2y} \\ V_{2x} & V_{2y} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} aV_{1x} + cV_{1y} & bV_{1x} + dV_{1y} \\ aV_{2x} + cV_{2y} & bV_{2x} + dV_{2y} \end{bmatrix} = \begin{bmatrix} \bar{V}_1^* \\ \bar{V}_2^* \end{bmatrix}. \quad (2-44)$$

Векторным произведением векторов \bar{V}_1^* и \bar{V}_2^* будет

$$\bar{V}_1^* \times \bar{V}_2^* = (ad - cb)(V_{1x}V_{2y} - V_{2x}V_{1y})\bar{k} = |\bar{V}_1^*| |\bar{V}_2^*| \bar{k} \sin \theta. \quad (2-45)$$

Аналогично, скалярное произведение будет равно

$$\begin{aligned} \bar{V}_1^* \cdot \bar{V}_2^* &= (a^2 + b^2)V_{1x}V_{2x} + (c^2 + d^2)V_{1y}V_{2y} + (ac + bd)(V_{1x}V_{2y} + V_{1y}V_{2x}) = \\ &= |\bar{V}_1^*| |\bar{V}_2^*| \cos \theta. \end{aligned} \quad (2-46)$$

Требуется, чтобы значения векторов, как и угол между ними, оставались постоянными. Сравнивая уравнения (2-42), (2-46) и (2-43), (2-45), а также приравнивая коэффициенты подобных членов, получим

$$a^2 + b^2 = 1, \quad (2-47a)$$

$$c^2 + d^2 = 1, \quad (2-47b)$$

$$ac + bd = 0, \quad (2-47c)$$

$$ad - bc = +1. \quad (2-48)$$

Выражения (2-47a, b, c) соответствуют условиям ортогональности матрицы, т. е.

$$[T][T]^{-1} = [T][T]^T = [I]$$

или

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Выражение (2-48) требует, чтобы определитель матрицы преобразования был равен +1.

Таким образом, при полном повороте углы между пересекающимися прямыми сохраняются. Данный результат распространяется также и на операцию отражения, ортогональная матрица которого имеет определитель, равный -1. В этом случае величины векторов сохраняются, но угол между преобразованными векторами в действительности равен $2\pi - \theta$. (Следовательно, в общем случае угол не сохраняется. Однако перпендикулярные прямые преобразуются в перпендикулярные прямые. Поскольку $\sin(2\pi - \theta) = -\sin \theta$, $ad - bc = -1$, полные повороты и отражения называются преобразованиями жесткой конструкции. Кроме того, несколько минут анализа или экспериментирования приводят к выводу, что равномерное масштабирование также сохраняет неизменным угол между пересекающимися прямыми, но не величину преобразуемых векторов. Поскольку ортогональная матрица сохраняет угол между векторами и их величины, матрица однородного масштабирования не является ортогональной.)

2-15 ПЕРЕМЕЩЕНИЯ И ОДНОРОДНЫЕ КООРДИНАТЫ

В предыдущих разделах был рассмотрен ряд преобразований, совершаемых с помощью (2×2) -матрицы общего преобразования. Среди них поворот, отражение, масштабирование, сдвиг и другие. Ранее отмечалось, что исходная система координат инвариантна по отношению ко всем перечисленным преобразованиям. Однако возникает необходимость изменять положение начала координат, т. е. преобразовывать каждую точку на плоскости. Этого можно достичь путем перемещения точки начала координат или любой другой точки на плоскости

$$x^* = ax + cy + m,$$

$$y^* = bx + dy + n.$$

К сожалению, нельзя ввести константы перемещения m и n в (2×2) -матрицу преобразования, так как это не пространство!

Данное затруднение можно преодолеть, используя однородные координаты. Однородные координаты неоднородного координатного вектора $[x \ y]$ представляют собой тройку $[x' \ y' \ h]$, где $x = x'/h$, $y = y'/h$, а h – некоторое вещественное число. Заметим, что случай $h = 0$ является особым. Всегда существует один набор однородных координат вида $[x \ y \ 1]$. Мы выбрали эту форму, чтобы представить координатный вектор $[x \ y]$ на физической плоскости xy . Все остальные однородные координаты представляются в виде $[hx \ hy \ h]$. Данные координаты не сохраняют однозначности, например, все следующие координаты $[6 \ 4 \ 2]$, $[12 \ 8 \ 4]$, $[3 \ 2 \ 1]$ представляют физическую точку $(3,2)$.

Матрица преобразования для однородных координат имеет размер 3×3 . В частности,

$$[T] = \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ m & n & 1 \end{bmatrix}, \quad (2-49)$$

где действие элементов a , b , c и d верхней части (2×2) -матрицы точно соответствует действиям, рассмотренным ранее. Элементы m и n являются коэффициентами перемещения в направлениях x и y соответственно. Полная двумерная матрица преобразования имеет вид

$$[x^* \ y^* \ 1] = [x \ y \ 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & 1 \end{bmatrix} = [x + m \ y + n \ 1]. \quad (2-50)$$

Отметим, что каждая точка плоскости и даже начало координат $x = y = 0$ теперь могут быть преобразованы.

2-16 ПОВОРОТ ВОКРУГ ПРОИЗВОЛЬНОЙ ТОЧКИ

Ранее мы рассматривали вращение, совершающееся вокруг начала координат. Однородные координаты предусматривают механизм выполнения поворотов вокруг точек, отличных от начала координат. В общем случае поворот вокруг произвольной точки может быть реализован посредством ее перемещения в начало координат, выполнения требуемого поворота и последующего перемещения результата обратно в исходный центр вращения. Таким образом, поворот вектора $[x \ y \ 1]$ вокруг точки m, n на произвольный угол можно осуществить следующим образом:

$$[x^* \ y^* \ 1] = [x \ y \ 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -m & -n & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & 1 \end{bmatrix}. \quad (2-51)$$

Выполняя действия над двумя внутренними матрицами, можно записать

$$[x^* \ y^* \ 1] = [x \ y \ 1] \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ \{-m(\cos \theta - 1)\} & \{-n(\cos \theta - 1)\} & 1 \\ \{+n \sin \theta\} & \{-m \sin \theta\} & 1 \end{bmatrix}. \quad (2-52)$$

Рассмотрим пример, иллюстрирующий данный результат.

Пример 2-6 Поворот относительно произвольной точки

Предположим, что центр объекта лежит в точке $[4 \ 3]$. Требуется повернуть объект на прямой угол против часовой стрелки вокруг центра. Используя матрицу

$$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

проведем поворот вокруг начала координат, не совпадающего с центром объекта. Обязательной процедурой преобразования является прежде всего такое перемещение объекта, чтобы желаемый центр вращения оказался в начале координат. Это достигается с помощью следующей матрицы перемещения:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & -3 & 1 \end{bmatrix}.$$

Далее применяем матрицу поворота и наконец с помощью матрицы перемещения приведем результаты поворота обратно к первоначальному центру. Вся операция

$$[x^* \ y^* \ 1] = [x \ y \ 1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & -3 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix}$$

может быть реализована одной матрицей путем простого перемножения отдельных матриц, т. е.

$$[x^* \ y^* \ 1] = [x \ y \ 1] = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 7 & -1 & 1 \end{bmatrix}.$$

2-17 ОТРАЖЕНИЕ ОТНОСИТЕЛЬНО ПРОИЗВОЛЬНОЙ ПРЯМОЙ

Ранее в разд. 2-10 обсуждалось отражение относительно прямых, проходящих через начало координат. Иногда требуется выполнить отражение объекта относительно прямой, не проходящей через точку начала координат. Это можно сделать, воспользовавшись процедурой, аналогичной вращению вокруг произвольной точки. Конкретно выполняются следующие действия:

- перемещение линии и объекта таким образом, чтобы линия прошла через начало координат;
- поворот линии и объекта вокруг точки начала координат до совпадения с одной из координатных осей;
- отражение относительно координатной оси;
- обратный поворот вокруг начала координат;
- перемещение в исходное положение.

В матричном виде данное преобразование имеет представление

$$[T] = [T'][R][R'][R]^{-1}[T']^{-1}, \quad (2-53)$$

где T' — матрица перемещения, R — матрица поворота вокруг начала координат, R' — матрица отражения.

Перемещения, повороты и отражения также применяются для преобразования произвольных фигур. Рассмотрим следующий пример.

Пример 2-7 Отражение относительно произвольной прямой

Рассмотрим прямую L и треугольник ABC (рис. 2-13a). Уравнение прямой L имеет вид

$$y = \frac{1}{2}(x + 4).$$

Координатные векторы $[2 \ 4 \ 1]$, $[4 \ 6 \ 1]$ и $[2 \ 6 \ 1]$ задают вершины треугольника ABC .

Прямая L пройдет через начало координат при перемещении ее на -2 единицы в направлении оси y . В результате этого при повороте вокруг начала координат на $-\operatorname{tg}^{-1}(1/2) = -26.57^\circ$ прямая совпадет с осью x . Выражение (2-33) используется для отражения треугольника относительно оси x , затем преобразованные координатные векторы треугольника поворачиваются и перемещаются к исходной ориентации. Комбинация преобразований будет иметь вид

$$\begin{aligned} [T] &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} & 0 \\ 1/\sqrt{5} & 2/\sqrt{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \\ &\times \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} & 0 \\ -1/\sqrt{5} & 2/\sqrt{5} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \\ [T] &= \begin{bmatrix} 3/5 & 4/5 & 0 \\ 4/5 & -3/5 & 0 \\ -8/5 & 16/5 & 1 \end{bmatrix}, \end{aligned}$$

и конкретно для координатных векторов треугольника $A^*B^*C^*$ имеем

$$\begin{bmatrix} 2 & 4 & 1 \\ 4 & 6 & 1 \\ 2 & 6 & 1 \end{bmatrix} \begin{bmatrix} 3/5 & 4/5 & 0 \\ 4/5 & -3/5 & 0 \\ -8/5 & 16/5 & 1 \end{bmatrix} = \begin{bmatrix} 14/5 & 12/5 & 1 \\ 28/5 & 14/5 & 1 \\ 22/5 & 6/5 & 1 \end{bmatrix}$$

(рис. 2-13a). Рис. 2-13 b, c, d, e иллюстрируют различные этапы данного преобразования.

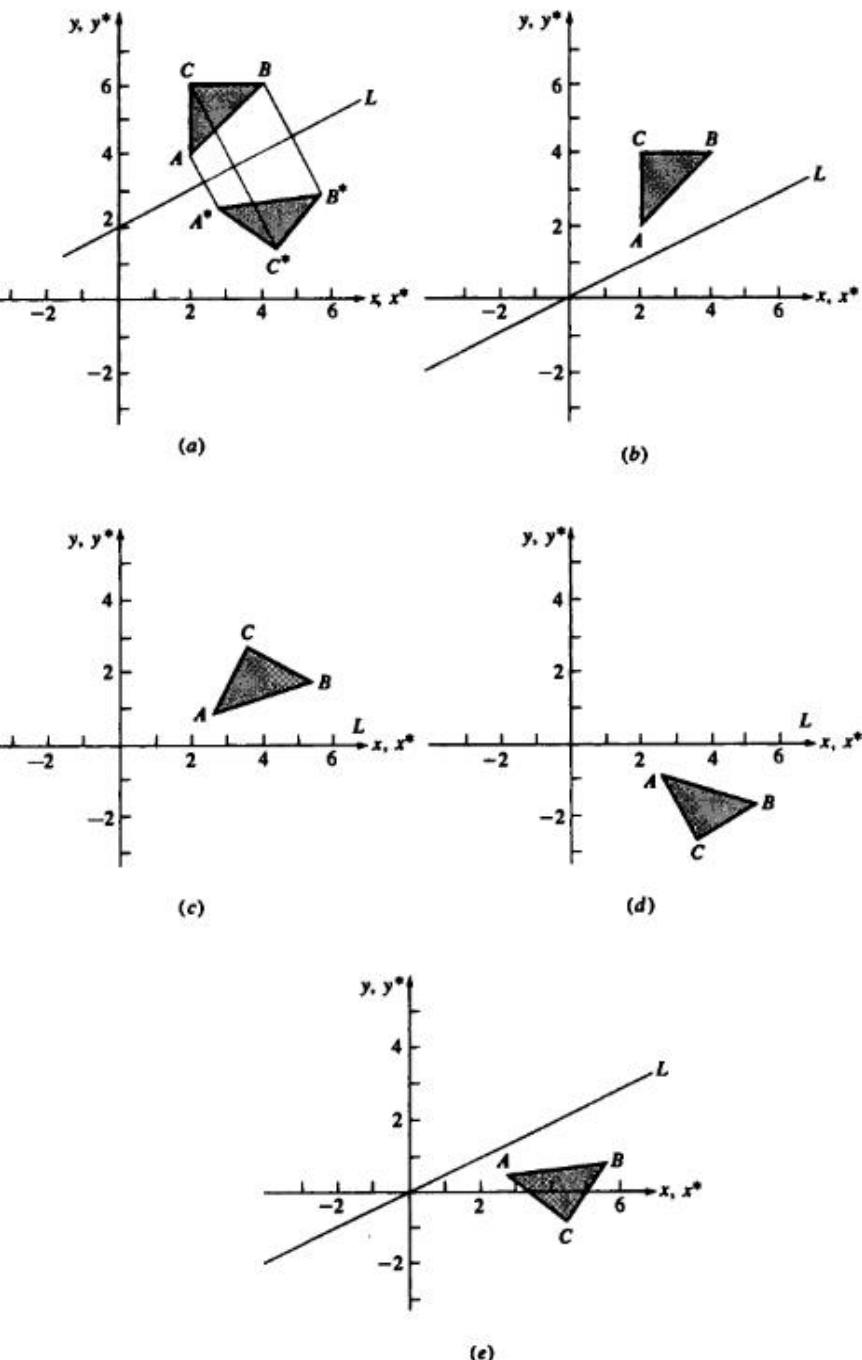


Рис. 2-13 Отражение относительно произвольной прямой: a) исходное и конечное положение; b) перенос прямой в начало координат; c) поворот до совпадения с осью x ; d) отражение относительно оси x ; e) обратный поворот; a) обратный перенос.

2-18 ПРОЕЦИРОВАНИЕ — ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ОДНОРОДНЫХ КООРДИНАТ

Матрицу преобразования размером 3×3 для двумерных однородных координат можно разбить на четыре части

$$[T] = \begin{bmatrix} a & b & : & p \\ c & d & : & q \\ \dots & \dots & \dots & \dots \\ m & n & : & s \end{bmatrix}. \quad (2-54)$$

Напомним, что a, b, c и d — коэффициенты масштабирования, вращения, отражения и сдвига соответственно. Элементы m и n задают перемещение. В двух предыдущих разделах коэффициенты имели значения $p = q = 0$ и $s = 1$. Установим величины p и q не равными 0. Какой эффект мы получим? В данном случае полезно рассмотреть геометрическую интерпретацию.

При $p = q = 0$ и $s = 1$ однородные координаты преобразованных векторов всегда равны $h = 1$. Геометрически данный результат интерпретируется как ограничение преобразования физической плоскостью $h = 1$.

Для иллюстрации эффекта преобразования при p и q , отличных от нуля, рассмотрим следующее выражение:

$$\begin{bmatrix} X & Y & h \end{bmatrix} = \begin{bmatrix} hx & hy & h \end{bmatrix} = \begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & p \\ 0 & 1 & q \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x & y & (px + qy + 1) \end{bmatrix}. \quad (2-55)$$

Здесь $X = hx$, $Y = hy$ и $h = px + qy + 1$. Преобразованный координатный вектор, выраженный в однородных координатах, лежит теперь в трехмерном пространстве, определенном как $h = px + qy + 1$. Это преобразование показано на рис. 2-14, где отрезок AB , принадлежащий физической плоскости $h = 1$, преобразуется в CD со значением $h \neq 1$, т. е. $pX + qY - h + 1 = 0$.

Однако представляют интерес результаты, принадлежащие физической плоскости с $h = 1$, которые можно получить путем геометрического проецирования прямой CD с плоскости $h \neq 1$ обратно на плоскость $h = 1$ с использованием для этого проецирующих лучей, проходящих через начало координат. Из рис. 2-14, используя правило подобия треугольников, получим

$$x^* = \frac{X}{h} \quad y^* = \frac{Y}{h}$$

или в однородных координатах

$$\begin{bmatrix} x^* & y^* & 1 \end{bmatrix} = \begin{bmatrix} Y & Y & 1 \end{bmatrix}.$$

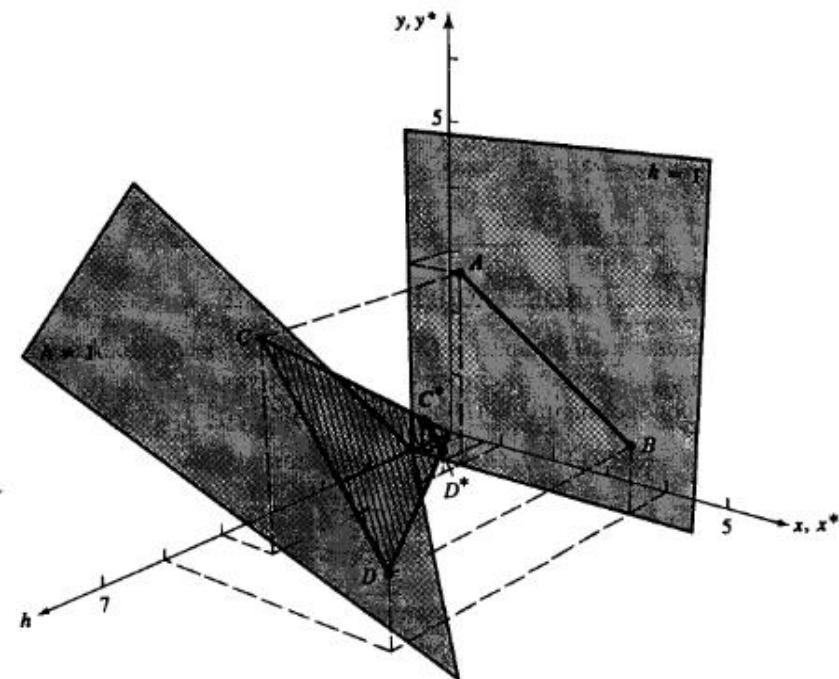


Рис. 2-14 Преобразование из физической плоскости ($h = 1$) на плоскость $h \neq 1$ и проецирование обратно на физическую плоскость.

После этого, нормализуя выражение (2-55) делением однородных координат на величину h , получаем

$$\begin{bmatrix} x^* & y^* & 1 \end{bmatrix} = \begin{bmatrix} X & Y & 1 \end{bmatrix} = \begin{bmatrix} \frac{x}{h} & \frac{y}{h} & 1 \end{bmatrix} = \begin{bmatrix} \frac{x}{px + qy + 1} & \frac{y}{px + qy + 1} & 1 \end{bmatrix} \quad (2-56)$$

или

$$x^* = \frac{X}{h} = \frac{x}{px + qy + 1}, \quad (2-57a)$$

$$y^* = \frac{Y}{h} = \frac{y}{px + qy + 1}. \quad (2-57b)$$

Детально действие преобразования рассмотрим на следующем примере.

Пример 2-8 Проецирование в однородных координатах

Для отрезка AB из рис. 2-14 имеем $p = q = 1$, $[A] = [1 \ 3 \ 1]$ и $[B] = [4 \ 1 \ 1]$,

$$\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} [T] = \begin{bmatrix} 1 & 3 & 1 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 5 \\ 4 & 1 & 6 \end{bmatrix}.$$

Таким образом, $[C] = [1 \ 3 \ 5]$ и $[D] = [4 \ 1 \ 6]$ на плоскости $h=x+y+1$. Проецируя обратно на плоскость $h=1$ путем деления на коэффициент однородных координат, проведем двумерное преобразование точек

$$[C^*] = [1 \ 3 \ 5] = [1/5 \ 3/5 \ 1],$$

$$[D^*] = [4 \ 1 \ 6] = [2/3 \ 1/6 \ 1].$$

Результат показан на рис. 2-14.

2-19 ПРОПОРЦИОНАЛЬНОЕ МАСШТАБИРОВАНИЕ

Оставшийся необъясненным элемент s (3×3)-матрицы преобразования соответствует пропорциональному масштабированию, при котором все компоненты вектора изменяются пропорционально. Покажем это, рассмотрев следующее преобразование:

$$[X \ Y \ h] = [x \ y \ 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{bmatrix} = [x \ y \ s]. \quad (2-58)$$

где $X = x$, $Y = y$ и $h = s$. После нормализации получим $X^* = x/s$ и $Y^* = y/s$. Таким образом, преобразование $[x \ y \ 1][T] = [x/s \ y/s \ 1]$ является равномерным масштабированием координатного вектора. Если $s < 1$, то происходит растяжение, а если $s > 1$ — сжатие.

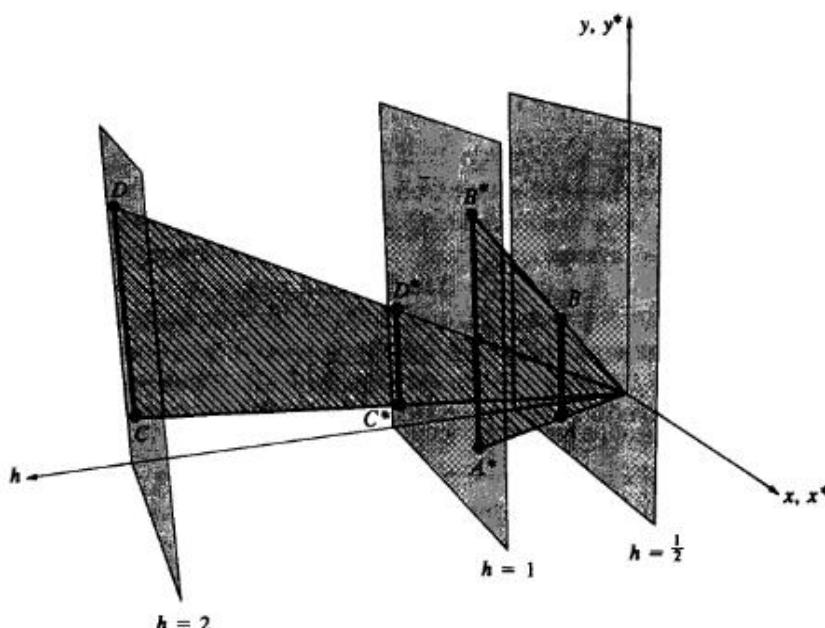


Рис. 2-15 Геометрическая интерпретация пропорционального масштабирования.

Заметим, что это преобразование осуществляется также в плоскости $h = 1$. Здесь $h = s = \text{const}$, и поэтому плоскость $h \neq 1$ параллельна плоскости $h = 1$. Геометрическая интерпретация данного эффекта показана на рис. 2-15. Если $s < 1$, то $h = \text{const}$ задает плоскость, лежащую между плоскостями $h = 1$ и $h = 0$. Следовательно, когда преобразуемая прямая AB проецируется обратно на плоскость $h = 1$, то A^*B^* увеличивается. Аналогично, если $s > 1$, то $h = \text{const}$ определяет плоскость, расположенную за плоскостью $h = 1$ и проходящую вдоль оси h . В случае проецирования прямой CD на плоскость $h = 1$ происходит уменьшение прямой C^*D^* .

2-20 ТОЧКИ БЕСКОНЕЧНОСТИ

Однородные координаты предоставляют удобный и эффективный способ нанесения точек из одной системы координат в соответствующие точки альтернативной координатной системы. Бесконечная область в одной координатной системе часто преобразуется в конечную область в альтернативной системе. При некорректном выборе переноса параллельность прямых может не сохраняться. Однако точки пересечения после преобразования оказываются снова в точках пересечения. Данное свойство используется для определения однородных координат представления точек бесконечности.

Рассмотрим пару пересекающихся прямых, заданных уравнениями

$$x + y = 1,$$

$$2x - 3y = 0.$$

Прямые пересекаются в точке с координатами $x = 3/5$, $y = 2/5$. Запишем уравнения в виде $x + y - 1 = 0$, $2x - 3y = 0$ и представим их в матричной форме

$$[x \ y \ 1] \begin{bmatrix} 1 & 2 \\ 1 & -3 \\ -1 & 0 \end{bmatrix} = [0 \ 0]$$

или

$$[X][M'] = [R].$$

Если матрица $[M']$ квадратная, то пересечение может быть получено путем обращения матрицы. Изменим систему исходных уравнений следующим образом:

$$x + y - 1 = 0,$$

$$2x - 3y = 0,$$

$$1 = 1,$$

или в матричной форме

$$[X][M] = [R],$$

т. е.

$$[x \ y \ 1] \begin{bmatrix} 1 & 2 & 0 \\ 1 & -3 & 0 \\ -1 & 0 & 1 \end{bmatrix} = [0 \ 0 \ 1].$$

Квадратная матрица, обратная данной [2-1], имеет следующий вид:

$$[M]^{-1} = \begin{bmatrix} 3/5 & 2/5 & 0 \\ 1/5 & -1/5 & 0 \\ 3/5 & 2/5 & 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 & 2 & 0 \\ 1 & -1 & 0 \\ 3 & 2 & 5 \end{bmatrix}.$$

Умножая обе части уравнения на $[M]^{-1}$ и учитывая, что $[M][M]^{-1} = [I]$ является тождественной матрицей, получим

$$[x \ y \ 1] = \frac{1}{5} [0 \ 0 \ 1] \begin{bmatrix} 3 & 2 & 0 \\ 1 & -1 & 0 \\ 3 & 2 & 5 \end{bmatrix} = [3/5 \ 2/5 \ 1].$$

Таким образом, точка пересечения опять имеет координаты $x = 3/5$, $y = 2/5$.

Рассмотрим теперь две параллельные прямые, заданные следующим образом:

$$\begin{aligned} x + y &= 1, \\ x + y &= 0. \end{aligned}$$

По определению геометрии Евклида, точка пересечения двух параллельных прямых расположена в бесконечности. Продолжая предыдущие рассуждения, вычислим точку пересечения этих прямых, заданных в матричной форме,

$$[x \ y \ 1] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = [0 \ 0 \ 1].$$

Однако несмотря на то что матрица квадратная, она не имеет обратной, так как две ее строки тождественны. Такая матрица называется сингулярной. Возможна иная формулировка с обратимой матрицей. Получим ее, переписывая систему уравнений следующим образом:

$$\begin{aligned} x + y - 1 &= 0, \\ x + y &= 0, \\ x &= x, \end{aligned}$$

или в матричной форме

$$[x \ y \ 1] \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} = [0 \ 0 \ x].$$

Таблица 2-1 Однородные координаты для точки [4 3]

h	x^*	y^*	X	Y
1	4	3	4	3
1/2	8	6	4	3
1/3	12	9	4	3
:				
1/10	40	30	4	3
:				
1/100	400	300	4	3
:				

В данном случае матрица не является сингулярной и существует обратная ей

$$[M]^{-1} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}.$$

Умножая обе части выражения на обратную матрицу, получаем

$$[x \ y \ 1] = [0 \ 0 \ x] \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & -1 & u \end{bmatrix} = [x \ -x \ 0] = x [1 \ -1 \ 0].$$

Результирующие однородные координаты $[1 \ -1 \ 0]$ определяют точку пересечения двух параллельных прямых, т. е. точку бесконечности. В частности, они представляют данную точку в направлении $[1 \ -1]$ двумерного пространства. В общем виде двумерный координатный вектор $[a \ b \ 0]$ представляет точку бесконечности на прямой $ay - bx = 0$. Приведем несколько примеров:

- $[1 \ 0 \ 0]$ точка на положительной оси x ,
- $[-1 \ 0 \ 0]$ точка на отрицательной оси x ,
- $[0 \ 1 \ 0]$ точка на положительной оси y ,
- $[0 \ -1 \ 0]$ точка на отрицательной оси y ,
- $[1 \ 1 \ 0]$ вдоль прямой $y = x$ в направлении $[1 \ 1]$.

Вектор с однородной компонентой $h = 0$ действительно представляет точку бесконечности и может быть также интерпретирован как движение к пределу (табл. 2-1).

Рассмотрим прямую $y^* = (3/4)x^*$ и точку $[X \ Y \ h] = [4 \ 3 \ 1]$. Напомним, что в однородных координатах не существует единственного представления координатного вектора (табл. 2-1). Точка $[4 \ 3 \ 1]$ представлена в однородных координатах по всем направлениям. Заметим, что в этой таблице при $h \rightarrow 0$ отношение y^*/x^* остается равным $3/4$, как и требуется для сохранения уравнения. Кроме этого, обратим внимание на то, что следующая пара $(x^* \ y^*)$, все точки которой располагаются на линии $y^* = (3/4)x^*$, быстро приближается к бесконечности. Таким

образом, предел при $h \rightarrow 0$ есть точка бесконечности, заданная в однородных координатах как $[X \ Y \ h] = [4 \ 3 \ 0]$.

Обратившись снова к рис. 2-15, легко продемонстрировать геометрическую интерпретацию процесса движения к пределу при $h \rightarrow 0$. Рассмотрим отрезок единичной длины, проходящий от точки начала координат в направлении $[1 \ 0]$ на плоскости $h = s$ ($s < 1$). При $s \rightarrow 0$ проекция этой прямой обратно на физическую плоскость $h = 1$ в направлении лучей, проходящих через начало координат, становится бесконечной длины. Следовательно, конечная точка прямой должна представляться точкой бесконечности на оси X .

2-21 ПРАВИЛА ВЫПОЛНЕНИЯ ПРЕОБРАЗОВАНИЙ

Для представления данных и выполнения преобразований с помощью умножения матриц используются различные соглашения. Наибольшее внимание нужно уделять формулировке задач и интерпретации результатов. Например, перед выполнением поворота необходимо получить ответы на следующие вопросы.

В правосторонней или левосторонней системе координат определяются поворачиваемые координатные векторы?

Вращается объект или система координат?

Как определяются положительный и отрицательный повороты?

Координаты записываются в виде строки или столбца матрицы?

Вокруг какой линии или оси осуществляется поворот?

В данном изложении используется правосторонняя система координат, объект вращается в неподвижной координатной системе, положительный поворот определяется правилом правой руки, т. е. поворот по часовой стрелке осуществляется вокруг оси при наблюдении от начала вдоль положительной оси, и координатные векторы представляются в виде строки матрицы.

Выражение (2-29) задает преобразование для положительного поворота вокруг начала координат или оси z . Так как вектор задается строкой матрицы, то матрицу преобразования следует разместить после данных или матрицы координатных векторов. Это преобразование задается путем умножения справа. В случае однородных координат для положительного поворота объекта на угол θ вокруг начала координат (оси z) использование умножения справа приводит к следующему результату:

$$[X^*] = [X][R],$$

$$[x^* \ y^* \ 1] = [x \ y \ 1] \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2-59)$$

Если мы подставим координатные векторы, заданные в однородных координатах в виде столбца матрицы, то поворот можно выполнить следующим образом:

$$[X^*] = [R]^{-1}[X],$$

2-21 Правила выполнения преобразований

$$\begin{bmatrix} x^* \\ y^* \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (2-60)$$

Выражение (2-60) называется преобразованием с умножением слева, так как матрица преобразования расположена перед столбцом координатного вектора или данных.

Заметим, что (3×3) -матрица в выражении (2-60) есть транспозиция (3×3) -матрицы из выражения (2-59). Это свидетельствует о независимости строк и столбцов матрицы.

Для того чтобы повернуть систему координат и оставить неизмененными координатные векторы, необходимо в выражении (2-59) заменить θ на $-\theta$. Вспомним, что $\sin \theta = -\sin(-\theta)$, а $\cos \theta = \cos(-\theta)$.

Теперь выражение (2-59) будет иметь вид

$$[x^* \ y^* \ 1] = [x \ y \ 1] = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2-61)$$

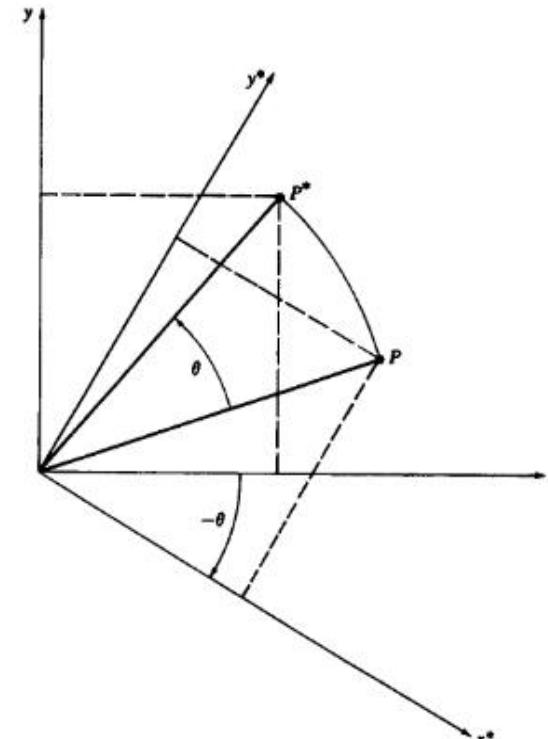


Рис. 2-16 Эквивалентность преобразования координатных векторов и систем координат.

Заметим, что (3×3) -матрица опять имеет обратную и также транспонируется в матрицу из (2-59).

Если вращается система координат и используется левосторонняя координатная система, то замену θ на $-\theta$ надо производить *дважды*, а уравнение (2-59) снова оказывается справедливым при допущении, что применяется последующее умножение на строку матрицы данных.

Заметим, что, как показано на рис. 2-16, вращение против часовой стрелки векторов, задающих объект, идентично повороту в том же направлении координатных осей при неподвижном объекте. Опять нет необходимости в изменении содержимого матрицы преобразования 3×3 , если нет других причин для ее редактирования. Эти несколько примеров показывают, насколько аккуратно необходимо выполнять матричные преобразования.

2-22 ЛИТЕРАТУРА

- 2-1 Fox L. *An Introduction to Numerical Linear Algebra*, Oxford University Press, London, 1964.
- 2-2 Forrest A. R. Coordinates, Transformations, and Visualization Techniques, CAD Group Document №. 23, Cambridge University, June 1969.

Глава 3

ПРОСТРАНСТВЕННЫЕ ПРЕОБРАЗОВАНИЯ И ПРОЕКЦИИ

3-1 ВВЕДЕНИЕ

Способность визуализировать или изображать пространственный объект является основой для понимания формы этого объекта. Кроме того, во многих случаях для этого важна способность вращать, переносить и строить виды проекций объекта. Все это легко демонстрируется на примере нашего знакомства с относительно сложным незнакомым объектом. Чтобы понять его форму, мы тут же начинаем вращать объект, отодвигать на расстояние вытянутой руки, передвигать вверх и вниз, вперед и назад и т. д. Чтобы сделать то же самое с помощью компьютера, мы должны распространить наш предшествующий двумерный анализ на три измерения. Основываясь на полученном опыте, мы немедленно вводим однородные координаты. Таким образом, точка в трехмерном пространстве $[x\ y\ z]$ представляется четырехмерным вектором

$$[x' \ y' \ z' \ h] = [x \ y \ z \ 1][T],$$

где $[T]$ является матрицей некоторого преобразования. Как и ранее, преобразование из однородных координат в обычные задается формулой

$$[x^* \ y^* \ z^* \ 1] = \left[\begin{array}{cccc} x' & y' & z' & 1 \\ \hline h & h & h & 1 \end{array} \right]. \quad (3-1)$$

Обобщенную матрицу преобразования размерности 4×4 для трехмерных однородных координат можно представить в следующем виде:

$$[T] = \left[\begin{array}{cccc} a & b & c & p \\ d & e & f & q \\ g & i & j & r \\ l & m & n & s \end{array} \right]. \quad (3-2)$$

Матрицу преобразования 4×4 из (3-2) можно разделить на четыре отдельные части:

$$\left[\begin{array}{ccc|c} & & & 3 \\ 3 \times 3 & & & \vdots \\ & & & \times \\ & & & \vdots \\ \cdots & \cdots & \cdots & \cdots \\ 1 \times 3 & & & \vdots \\ & & & 1 \times 1 \end{array} \right].$$

Верхняя левая (3×3) -подматрица задает линейное преобразование¹ в форме масштабирования, сдвига, отражения и вращения. Левая нижняя (1×3) -подматрица задает перемещение, а правая верхняя (3×1) -подматрица — перспективное преобразование. Последняя правая нижняя (1×1) -подматрица задает общее масштабирование. Общее преобразование, полученное после применения этой (4×4) -матрицы к однородному вектору и вычисления обычных координат, называется билинейным преобразованием². В общем случае данное преобразование осуществляет комбинацию сдвига, локального масштабирования, вращения, отражения, перемещения, перспективного преобразования и общего масштабирования.

3-2 ТРЕХМЕРНОЕ МАШТАБИРОВАНИЕ

Диагональные элементы (4×4) -матрицы обобщенного преобразования задают локальное и общее масштабирование. Для иллюстрации этого рассмотрим преобразование

$$\begin{aligned}[X][T] &= [x \ y \ z \ 1] \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & e & 0 & 0 \\ 0 & 0 & j & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= [ax \ ey \ jz \ 1] = [x^* \ y^* \ z^* \ 1], \end{aligned} \quad (3-3)$$

которое показывает действие локального масштабирования. Ниже приводится пример.

Пример 3-1 Локальное масштабирование

Рассмотрим прямоугольный параллелепипед (*RPP*), показанный на рис. 3-1 а со следующими однородными координатами вершин:

$$[X] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 2 & 3 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 3 & 0 & 1 \\ 0 & 3 & 0 & 1 \end{bmatrix}.$$

Чтобы получить единичный куб из *RPP* с помощью локального масштабирования, необходимы масштабные множители $1/2, 1/3, 1$ вдоль осей x, y, z соответственно. Преобразование локального масштабирования задается матрицей

$$[T] = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

¹Линейное преобразование — это преобразование, переводящее линейную комбинацию векторов в ту же самую линейную комбинацию преобразованных векторов.

²Билинейное преобразование — это результат двух последовательных линейных преобразований.

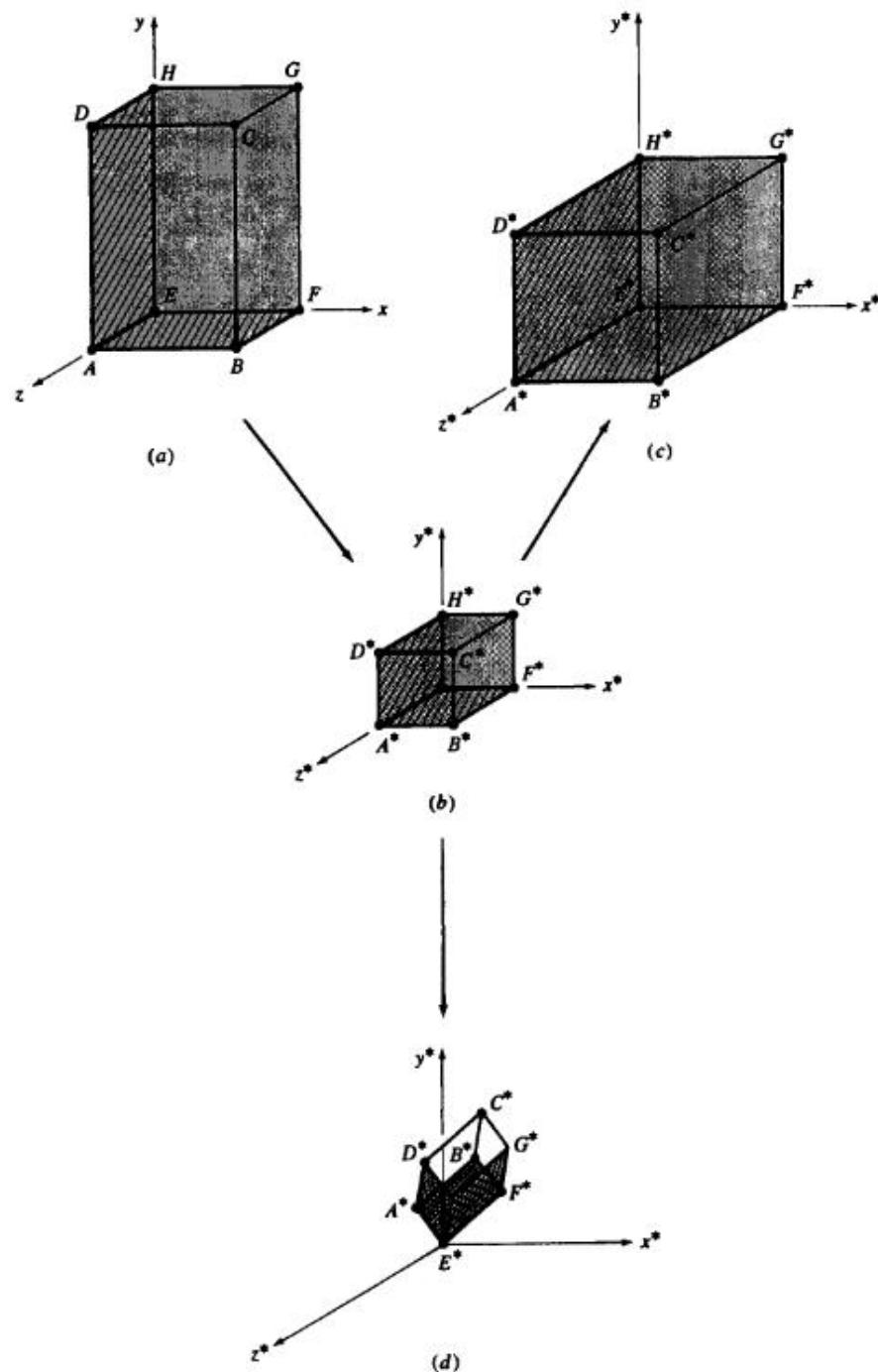


Рис. 3-1 Трехмерные масштабирования.

Результирующий куб имеет следующие однородные координаты вершин:

$$\begin{aligned}[X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 2 & 3 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 3 & 0 & 1 \\ 0 & 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}\end{aligned}$$

Заметим, что однородный координатный множитель h равен единице для каждой из преобразованных вершин. Результат масштабирования показан на рис. 3-1 б.

Общее масштабирование можно осуществить, воспользовавшись четвертым диагональным элементом, т. е.

$$[X][T] = [x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & s \end{bmatrix} = [x' \ y' \ z' \ s]. \quad (3-4)$$

Обычные или физические координаты имеют вид

$$[x^* \ y^* \ z^* \ 1] = \left[\frac{x'}{s} \ \frac{y'}{s} \ \frac{z'}{s} \ 1 \right].$$

Этот результат снова иллюстрируется на примере.

Пример 3-2 Общее масштабирование

Для общего масштабирования единичного куба, изображенного на рис. 3-1 б, на множитель два (удвоение размера), необходимо преобразование (см. (3-4))

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}.$$

Полученный в результате параллелепипед RPP имеет следующие однородные координаты вершин:

$$\begin{aligned}[X'] &= [X^*][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 0 & 1 & 0.5 \\ 1 & 0 & 1 & 0.5 \\ 1 & 1 & 1 & 0.5 \\ 0 & 1 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 1 & 0 & 0 & 0.5 \\ 1 & 1 & 0 & 0.5 \\ 0 & 1 & 0 & 0.5 \end{bmatrix}.\end{aligned}$$

Заметим, что однородный координатный множитель h для каждой из преобразованных вершин равен 0.5. Таким образом, для того чтобы получить обычные или физические координаты, каждый вектор необходимо разделить на h . Результат, показанный на рис. 3-1 с, равен

$$[X^*] = \begin{bmatrix} 0 & 0 & 2 & 1 \\ 2 & 0 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 0 & 2 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 2 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix}.$$

Заметим здесь, что, как и в случае двумерного общего масштабирования, однородный координатный множитель не равен единице. По аналогии с предыдущим обсуждением (см. разд. 2-18) это означает преобразование из физического объема $h = 1$ в другой объем в 4-мерном пространстве. Преобразованные физические координаты получаются проецированием через центр 4-мерной координатной системы обратно в физический объем $h = 1$. Как и ранее, если $s < 1$, происходит однородное расширение. Если $s > 1$, происходит однородное сжатие координатного вектора.

Такой же результат можно получить, используя одинаковые коэффициенты локальных масштабирований. В этом случае матрица преобразования имеет вид

$$[T] = \begin{bmatrix} 1/s & 0 & 0 & 0 \\ 0 & 1/s & 0 & 0 \\ 0 & 0 & 1/s & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Отметим, что здесь однородный координатный множитель равен единице, т. е. $h = 1$. Таким образом, все преобразование происходит в физическом объеме $h = 1$.

3-3 ТРЕХМЕРНЫЕ СДВИГИ

Недиагональные элементы в верхней левой 3×3 -подматрице обобщенной матрицы преобразования размером 4×4 задают сдвиги в трех измерениях, т. е.

$$\begin{aligned}[X][T] &= [x \ y \ z \ 1] \begin{bmatrix} 1 & b & c & 0 \\ d & 1 & f & 0 \\ g & i & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= [x + yd + gz \ bx + y + iz \ cx + fy + z \ 1].\end{aligned}\quad (3-5)$$

Поясним это на примере.

Пример 3-3 Сдвиги

Снова рассмотрим единичный куб, изображенный на рис. 3-1 б. Применив преобразование сдвигов

$$[T] = \begin{bmatrix} 1 & -0.85 & 0.25 & 0 \\ -0.75 & 1 & 0.7 & 0 \\ 0.5 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

получаем

$$\begin{aligned}[X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -0.85 & 0.25 & 0 \\ -0.75 & 1 & 0.7 & 0 \\ 0.5 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0.5 & 1 & 1 & 1 \\ 1.5 & 0.15 & 1.25 & 1 \\ 0.75 & 1.15 & 1.95 & 1 \\ -0.25 & 2 & 1.7 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & -0.85 & 0.25 & 1 \\ 0.25 & 0.15 & 0.95 & 1 \\ -0.75 & 1 & 0.7 & 1 \end{bmatrix}.\end{aligned}$$

Этот результат показан на рис. 3-1 д. Заметим, что во всех трех примерах остается неизменной вершина параллелепипеда, находящаяся в начале координат.

3-4 ТРЕХМЕРНОЕ ВРАЩЕНИЕ

Прежде чем переходить к трехмерному вращению вокруг произвольной оси, рассмотрим вращение вокруг каждой из координатных осей. При вращении вокруг оси x остаются неизменными x -координаты координатного вектора. Фактически

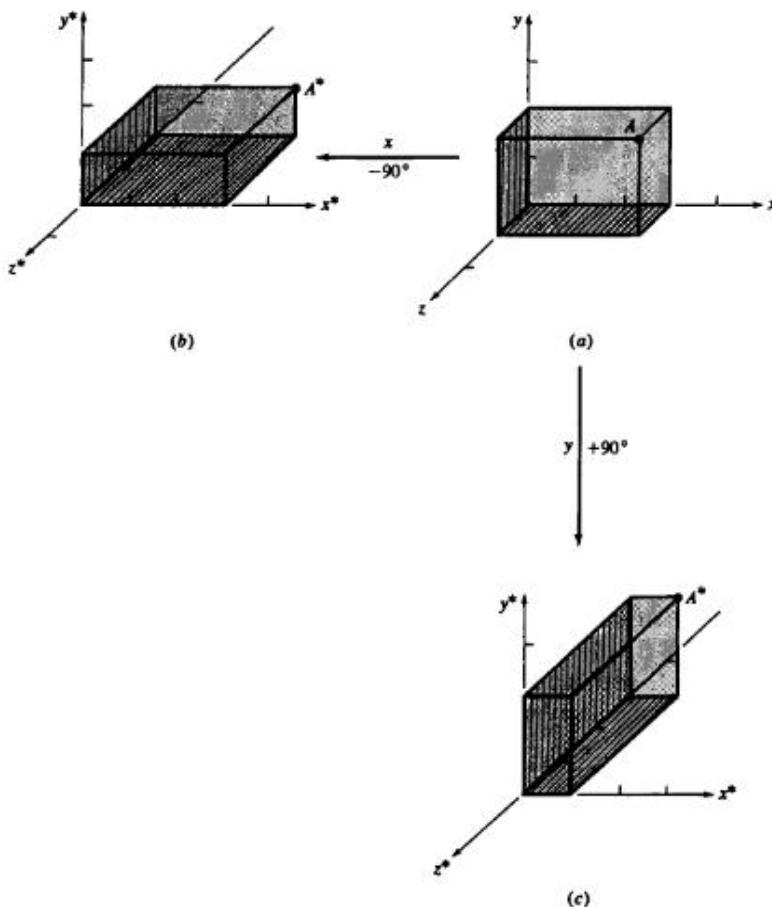


Рис. 3-2 Трехмерные повороты.

вращение происходит в плоскостях, перпендикулярных оси x . Аналогичным образом вращение вокруг осей y и z происходит в плоскостях, перпендикулярных осям y и z соответственно. Преобразование координатного вектора в каждой из этих плоскостей задается указанной в (2-29) матрицей двумерного вращения. Эта матрица и неизменность координаты x при вращении вокруг оси x позволяют записать 4×4 -преобразование однородных координат при повороте на угол θ в виде

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-6)$$

Вращение считается положительным в смысле правила правой руки, т. е. по часовой стрелке, если смотреть из начала координат в положительном направле-

нии оси вращения¹. На рис. 3-2 б показан параллелепипед, полученный поворотом на -90° вокруг оси x параллелепипеда с рис. 3-2 а.

Аналогично матрица преобразования для вращения вокруг оси z на угол ψ имеет вид

$$[T] = \begin{bmatrix} \cos \psi & \sin \psi & 0 & 0 \\ -\sin \psi & \cos \psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-7)$$

При вращении на угол ϕ вокруг оси y преобразование имеет вид

$$[T] = \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-8)$$

Заметим, что в (3-8) знаки у синусов противоположны знакам этих членов в равенствах (3-6) и (3-7). Это нужно для того, чтобы выполнялось соглашение о положительном направлении по правилу правой руки.

Из равенств (3-6)–(3-8) следует, что детерминант каждой из матриц преобразований равен +1, что и необходимо для чистого вращения. Более полно эти результаты проиллюстрирует пример.

Пример 3-4 Вращение

Рассмотрим прямоугольный параллелепипед, изображенный на рис. 3-2 а. Матрица $[X]$ координатного вектора имеет вид

$$[X] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 3 & 0 & 1 & 1 \\ 3 & 2 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 3 & 2 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} A$$

В данном случае строка, отмеченная буквой A в матрице $[X]$, соответствует точке A на рис. 3-2. Равенство (3-6) для поворота на $\theta = -90^\circ$ вокруг оси x приводит к следующему преобразованию:

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

¹ Правило правой руки для вращения формулируется следующим образом: направьте большей пальцем правой руки в положительном направлении оси вращения. Тогда естественное спиральное движение пальцев указывает положительное направление вращения.

Применение этого преобразования дает новые координаты:

$$\begin{aligned} [X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 3 & 0 & 1 & 1 \\ 3 & 2 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 3 & 2 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 1 & 0 & 1 \\ 3 & 1 & 0 & 1 \\ 3 & 1 & -2 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 3 & 0 & -2 & 1 \\ 0 & 0 & -2 & 1 \end{bmatrix} A^* \\ &= \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Отметим, что, как и должно быть, x -компоненты $[X]$ и $[X^*]$ идентичны. Результат данного поворота изображен на рис. 3-2 б.

Для поворота на угол $\phi = +90^\circ$ вокруг оси y уравнение (3-7) дает следующую матрицу преобразования:

$$[T'] = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Снова применив преобразование к исходному параллелепипеду, получим новые координаты:

$$\begin{aligned} [X^{**}] &= [X][T'] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 3 & 0 & 1 & 1 \\ 3 & 2 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 3 & 2 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & -3 & 1 \\ 1 & 2 & -3 & 1 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -3 & 1 \\ 0 & 2 & -3 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} A^* \\ &= \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & -3 & 1 \\ 1 & 2 & -3 & 1 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -3 & 1 \\ 0 & 2 & -3 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Заметим, что в этом случае идентичны y -компоненты $[X]$ и $[X^{**}]$. Результат изображен на рис. 3-2 с.

Так как трехмерные вращения получаются с помощью перемножения матриц, то они не коммутативны; т. е. порядок перемножения влияет на конечный результат.

тат (см. разд. 2-12). Чтобы показать это, рассмотрим два последовательных поворота на один и тот же угол — сначала вокруг оси x , затем вокруг оси y . Используя уравнения (3-6) и (3-8) с $\theta = \phi$, мы получим

$$\begin{aligned} [T] &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & 0 & -\sin \theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \theta & 0 & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \cos^2 \theta & 0 & -\sin \theta & 0 \\ \sin^2 \theta & \cos \theta & \cos \theta \sin \theta & 0 \\ \cos \theta \sin \theta & -\sin \theta & \cos^2 \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \quad (3-9)$$

С другой стороны, обратная операция, т. е. поворот вокруг оси y , а потом вокруг оси x с углом $\theta = \phi$ дает

$$\begin{aligned} [T] &= \begin{bmatrix} \cos \theta & 0 & -\sin \theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \theta & 0 & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \cos \theta & \sin^2 \theta & -\cos \theta \sin \theta & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ \sin \theta & -\cos \theta \sin \theta & \cos^2 \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \quad (3-10)$$

Сравнивая правые части (3-9) и (3-10), видим, что они не одинаковы. Если надо сделать более одного поворота, то следует помнить о некоммутируемости трехмерных вращений.

На рис. 3-3 с 3-3 d штриховой линией изображен результат преобразования, состоящего из двух поворотов на 90° при помощи произведения матриц из (3-9) для объекта, показанного на рис. 3-3 a. Осуществляя повороты, заданные (3-10), в обратном порядке, получим фигуры, нарисованные сплошными линиями на рис. 3-3 b и 3-3 d. Рис. 3-3 d наглядно показывает, что при изменении порядка вращения получаются различные результаты. Приведенный ниже численный пример иллюстрирует это.

Пример 3-5 Комбинированные повороты

Объект на рис. 3-2 a имеет следующие координатные векторы:

$$[X] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 2 & 3 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 3 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} A$$

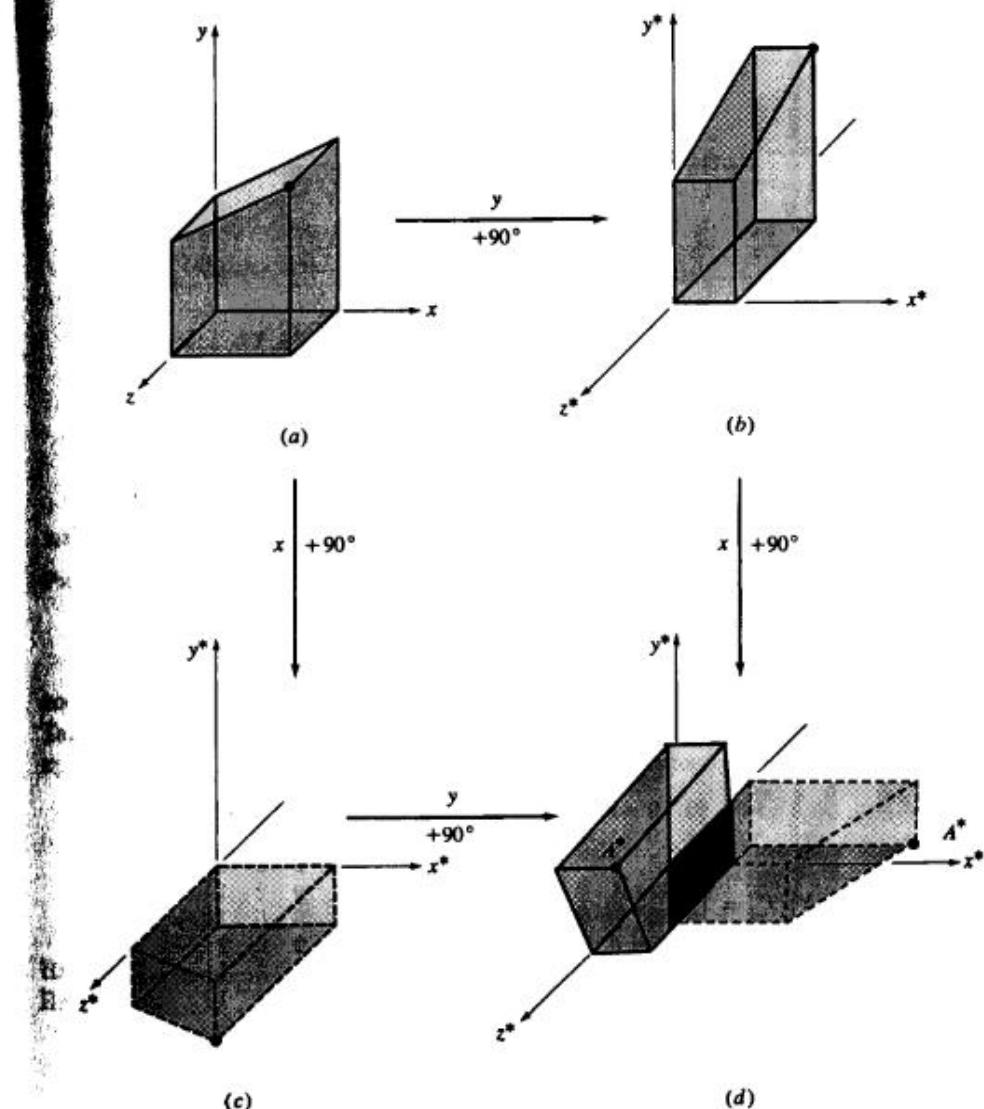


Рис. 3-3 Некоммутативность трехмерных поворотов.

Общая матрица для вращения сначала вокруг оси x на угол $\theta = 90^\circ$, а затем вокруг оси y на угол $\phi = 90^\circ$ задается уравнением (3-9) в виде

$$[T] = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразованные координатные векторы равны

$$\begin{aligned}[X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 2 & 3 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 3 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & -1 & 0 & 1 \\ 0 & -1 & -2 & 1 \\ 3 & -1 & -2 & 1 \\ 2 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & 1 \\ 3 & 0 & -2 & 1 \\ 2 & 0 & 0 & 1 \end{bmatrix} A^*.\end{aligned}$$

Преобразованный объект изображен штриховой линией на рис. 3-3 d.
Общая матрица для вращения сначала вокруг оси y на угол $\phi = 90^\circ$, а
затем вокруг оси x на угол $\theta = 90^\circ$ задается уравнением (3-10) в виде

$$[T'] = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

В этом случае преобразованные координатные векторы равны

$$\begin{aligned}[X^{**}] &= [X][T'] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 2 & 3 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 2 & 3 & 0 & 1 \\ 0 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 2 & 3 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 0 & 2 & 3 & 1 \\ 0 & 0 & 2 & 1 \end{bmatrix} A^{**}.\end{aligned}$$

Преобразованный объект изображен сплошными линиями на рис. 3-3 d.
Сравнение двух числовых результатов также ясно показывает, что ориентация преобразованных объектов совершенно различна. Следовательно, порядок перемножения матриц очень важен.

3-5 ТРЕХМЕРНОЕ ОТРАЖЕНИЕ

Некоторые ориентации трехмерного объекта нельзя получить одними вращениями, требуются преобразования отражения. В трехмерном пространстве отражение происходит относительно плоскости. По аналогии с обсуждавшимся ранее двумерным отражением (см. разд. 2-10), трехмерное отражение относительно плоскости эквивалентно вращению вокруг оси в трехмерном пространстве в четырехмерное пространство и обратно в исходное трехмерное пространство. Для чистого отражения детерминант матрицы равен -1 .

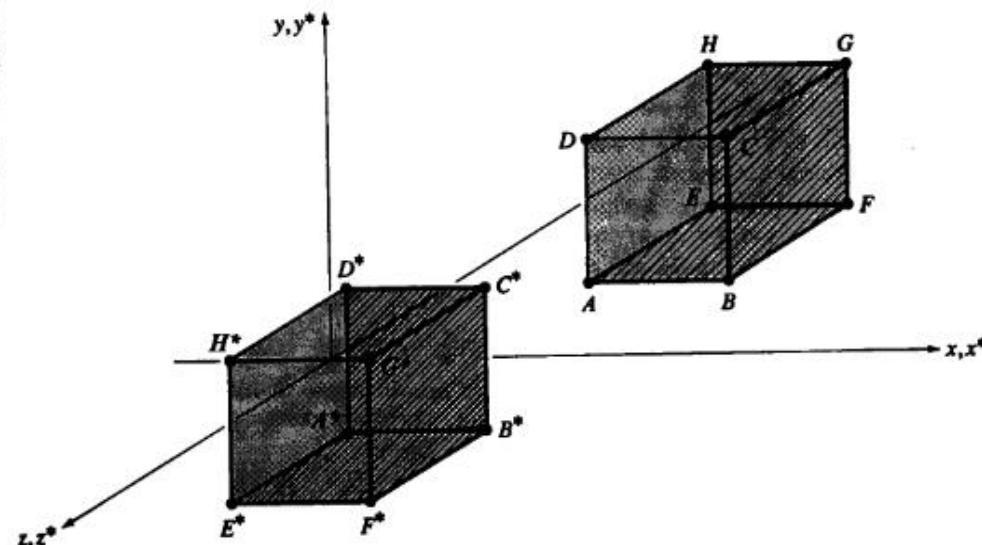


Рис. 3-4 Трехмерное отражение относительно плоскости xy .

При отражении относительно плоскости xy изменяются только значения z -координаты координатного вектора объекта. В самом деле, они изменяют знак. Таким образом, матрица преобразования для отражения относительно плоскости xy равна

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-11)$$

На рис. 3-4 изображено отражение единичного куба относительно плоскости xy . При отражении относительно плоскости yz

$$[T] = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (3-12)$$

а для отражения относительно плоскости xz

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-13)$$

Эти результаты иллюстрируются в приведенном ниже численном примере.

Пример 3-6 Отражение

Параллелепипед $ABCDEFGH$, изображенный на рис. 3-4, имеет координатный вектор

$$[X] = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 2 & 0 & -1 & 1 \\ 2 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 0 & -2 & 1 \\ 2 & 0 & -2 & 1 \\ 2 & 1 & -2 & 1 \\ 1 & 1 & -2 & 1 \end{bmatrix}.$$

Матрица преобразования для отражения относительно плоскости xy задается посредством (3-11). После отражения преобразованные координатные векторы равны

$$[X^*] = [X][T] = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 2 & 0 & -1 & 1 \\ 2 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 0 & -2 & 1 \\ 2 & 0 & -2 & 1 \\ 2 & 1 & -2 & 1 \\ 1 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 0 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 1 \\ 2 & 0 & 2 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix}.$$

Результат $A^*B^*C^*D^*E^*F^*G^*H^*$ показан на рис. 3-4.

3-6 ПРОСТРАНСТВЕННЫЙ ПЕРЕНОС

Матрица пространственного переноса имеет вид:

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ l & m & n & 1 \end{bmatrix}. \quad (3-14)$$

Перемещенные однородные координаты получаются с помощью преобразования

$$[x' \ y' \ z' \ h] = [x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ l & m & n & 1 \end{bmatrix}.$$

Выполнив умножение, получим

$$[x' \ y' \ z' \ h] = [(x+l) \ (y+m) \ (z+n) \ 1]. \quad (3-15)$$

Из этого следует, что преобразованные физические координаты равны

$$x^* = x + l,$$

$$y^* = y + m,$$

$$z^* = z + n.$$

3-7 КОМПОЗИЦИИ ПРЕОБРАЗОВАНИЙ

Последовательные преобразования могут быть скомбинированы или объединены в одно (4×4)-преобразование, дающее тот же самый результат. Так как перемежение матриц является некоммутативной операцией, то важен порядок ее выполнения (в общем случае $[A][B] \neq [B][A]$). Правильный порядок определяется положением конкретной матрицы преобразования относительно матрицы координатного вектора. Матрица, ближайшая к матрице координатного вектора, задает первое преобразование, а последняя — последнее преобразование. Математически это можно записать следующим образом:

$$[X][T] = [X][T_1][T_2][T_3][T_4] \dots,$$

где

$$[T] = [T_1][T_2][T_3][T_4] \dots$$

и $[T_i]$ являются произвольной комбинацией матриц масштабирования, сдвига, вращения, отражения, переноса, перспективного преобразования и проецирования. Так как перспективные преобразования искажают геометрические объекты (см. разд. 3-15), а преобразования проецирования приводят к потере информации (см. разд. 3-12), то в случае наличия этих матриц они должны быть расположены соответственно предпоследней и последней по порядку.

Эти идеи проиллюстрированы в приведенном ниже примере.

Пример 3-7 Композиции преобразований

Рассмотрим для заданного в однородных координатах координатного вектора $[1 \ 2 \ 1 \ 1]$ результат переноса в направлениях x, y, z на $-1, -1, -1$ соответственно, а затем поворота на $+30^\circ$ вокруг оси x и поворота на $+45^\circ$ вокруг оси y .

Сначала получим матрицу комбинированного преобразования. Из равенств (3-14), (3-6) и (3-8) следует, что

$$\begin{aligned} [T] &= [Tr][R_x][R_y] = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ l & m & n & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ l & m & n & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ \sin \phi \sin \theta & \cos \theta & \cos \phi \sin \theta & 0 \\ \sin \phi \cos \theta & -\sin \theta & \cos \phi \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ \sin \phi \sin \theta & \cos \theta & \cos \phi \sin \theta & 0 \\ \sin \phi \cos \theta & -\sin \theta & \cos \phi \cos \theta & 0 \\ l \cos \phi & m \cos \theta & -l \sin \phi & 1 \\ +m \sin \phi \sin \theta & -n \sin \theta & +m \cos \phi \sin \theta & 0 \\ +n \sin \phi \cos \theta & +n \cos \theta & +n \cos \phi \cos \theta & 0 \end{bmatrix}, \end{aligned} \quad (3-16)$$

где θ и ϕ — соответственно углы вращения относительно осей x и y ; а l, m, n — величины переноса в направлениях x, y, z .

В общем случае для координатного вектора мы имеем

$$\begin{aligned}[X][T] &= [x \ y \ z \ 1] \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ \sin \phi \sin \theta & \cos \theta & \cos \phi \sin \theta & 0 \\ \sin \phi \cos \theta & -\sin \theta & \cos \phi \cos \theta & 0 \\ l \cos \phi & m \cos \theta & -l \sin \phi & 1 \\ +m \sin \phi \sin \theta & -n \sin \theta & +m \cos \phi \sin \theta & 0 \\ +n \sin \phi \cos \theta & +n \cos \phi \cos \theta & +n \cos \phi \sin \theta & 1 \end{bmatrix} = \\ &= \begin{bmatrix} (x+l) \cos \phi & (y+m) \cos \theta & -(x+l) \sin \phi & 1 \\ +(y+m) \sin \phi \sin \theta & -(z+n) \sin \theta & +(y+m) \cos \phi \sin \theta & 0 \\ +(z+n) \sin \phi \cos \theta & +(z+n) \cos \phi \cos \theta & +(z+n) \cos \phi \sin \theta & 1 \end{bmatrix}.\end{aligned}$$

Для конкретных значений $\theta = +30^\circ$, $\phi = +45^\circ$, $l = -1$, $m = 1$, $n = -1$ преобразуемый координатный вектор имеет вид $[3 \ 2 \ 1 \ 1]$.

$$[X][T] = [3 \ 2 \ 1 \ 1] \begin{bmatrix} 0.707 & 0 & -0.707 & 0 \\ 0.354 & 0.866 & 0.354 & 0 \\ 0.612 & -0.5 & 0.612 & 0 \\ -1.673 & -0.366 & -0.259 & 1 \end{bmatrix},$$

$$[X][T] = [1.768 \ 0.866 \ -1.061 \ 1].$$

Чтобы убедиться, что общая матрица дает тот же самый результат, как и последовательное применение матриц, рассмотрим

$$\begin{aligned}[X'] &= [X][Tr] = \\ &= [3 \ 2 \ 1 \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 1 \end{bmatrix} = \\ &= [2 \ 1 \ 0 \ 1],\end{aligned}$$

$$\begin{aligned}[X''] &= [X'][R_x] = [2 \ 1 \ 0 \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.866 & 0.5 & 0 \\ 0 & -0.5 & 0.866 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= [2 \ 0.866 \ 0.5 \ 1],\end{aligned}$$

$$[X'''] = [X''][R_y] = [2 \ 0.866 \ 0.5 \ 1] \begin{bmatrix} 0.707 & 0 & -0.707 & 0 \\ 0 & 1 & 0 & 0 \\ 0.707 & 0 & 0.707 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$[X'''] = [1.768 \ 0.866 \ -1.061 \ 1],$$

что доказывает наше утверждение.

3-8 ПОВОРОТЫ ВОКРУГ ОСИ, ПАРАЛЛЕЛЬНОЙ КООРДИНАТНОЙ ОСИ

Преобразования, заданные равенствами (3-6)–(3-8), описывают вращение вокруг координатных осей x , y и z . Однако часто бывает необходимо вращать объект вокруг оси, не совпадающей с этими тремя. Мы рассмотрим частный случай для оси, параллельной одной из координатных осей x , y или z . На рис. 3-5 изображено тело в локальной системе осей $x'y'z'$, параллельных фиксированной глобальной системе xyz . Вращение тела вокруг любой из локальных осей x' , y' или z' выполняется с помощью следующей процедуры:

- переместить тело так, чтобы локальная ось совпала с координатной;
- повернуть вокруг указанной оси;
- переместить преобразованное тело в исходное положение.

Математически это можно записать так:

$$[X^*] = [X][Tr][R_x][Tr]^{-1},$$

где

- $[X^*]$ – преобразованное тело,
- $[X]$ – исходное тело,
- $[Tr]$ – матрица перемещения,
- $[R_x]$ – соответствующая матрица поворота,
- $[Tr]^{-1}$ – матрица, обратная к матрице перемещения.

Ниже приводится иллюстративный пример.

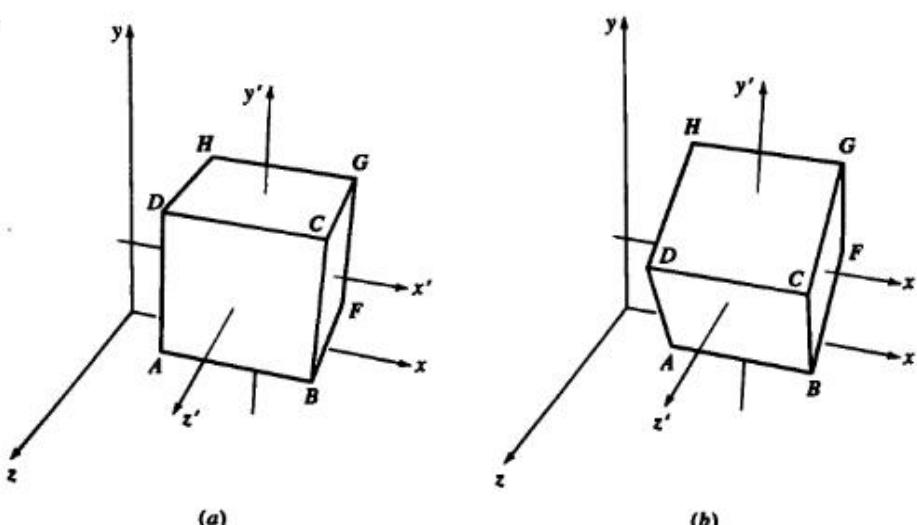


Рис. 3-5 Поворот вокруг оси, параллельной одной из координатных осей.

Пример 3-8 Единственное относительное вращение

Рассмотрим параллелепипед, изображенный на рис. 3-5 а, заданный координатными векторами

$$[X] = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{bmatrix} \begin{array}{l} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{array}$$

относительно глобальной системы координат xyz . Повернем параллелепипед на $\theta = +30^\circ$ относительно локальной оси x' , проходящей через центр параллелепипеда. Предполагается, что начало локальной системы координат находится в центре параллелепипеда. Координаты этого центра равны $[x_c \ y_c \ z_c \ 1] = [3/2 \ 3/2 \ 3/2 \ 1]$. Вращение осуществляется следующим образом:

$$[X^*] = [X] [Tr] [R_x] [Tr]^{-1},$$

где

$$[Tr] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -y_c & -z_c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -3/2 & -3/2 & 1 \end{bmatrix},$$

$$[R_x] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.866 & 0.5 & 0 \\ 0 & -0.5 & 0.866 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

и

$$[Tr]^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & y_c & z_c & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 3/2 & 3/2 & 1 \end{bmatrix}.$$

Первая матрица $[Tr]$ сдвигает параллелепипед параллельно плоскости $x = 0$ до тех пор, пока ось x' не совпадет с осью x . Вторая матрица $[R_x]$ выполняет требуемое вращение вокруг оси x , третья матрица $[Tr]^{-1}$ переносит ось x' , следовательно, и повернутый параллелепипед, обратно в исходное положение.

Объединив эти три матрицы, получим

$$\begin{aligned} [T] &= [Tr] [R_x] [Tr]^{-1} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & y_c(1 - \cos \theta) + z_c \sin \theta & z_c(1 - \cos \theta) - y_c \sin \theta & 1 \end{bmatrix}. \end{aligned}$$

3-8 Повороты вокруг оси, параллельной координатной оси

После подстановки числовых значений преобразованные координаты примут вид:

$$[X'] = [X] [T] = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.866 & 0.5 & 0 \\ 0 & -0.5 & 0.866 & 0 \\ 0 & 0.951 & -0.549 & 1 \end{bmatrix},$$

$$[X'] = \begin{bmatrix} 1 & 0.817 & 1.683 & 1 \\ 2 & 0.817 & 1.683 & 1 \\ 2 & 1.683 & 2.183 & 1 \\ 1 & 1.683 & 2.183 & 1 \\ 1 & 1.317 & 0.817 & 1 \\ 2 & 1.317 & 0.817 & 1 \\ 2 & 2.183 & 1.317 & 1 \\ 1 & 2.183 & 1.317 & 1 \end{bmatrix} \begin{array}{l} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \end{array}$$

Результат изображен на рис. 3-5 б.

В предыдущем примере требовалось только вращение вокруг единственной оси, параллельной одной из координатных осей. Таким образом надо только было сделать так, чтобы ось вращения совпала с соответствующей координатной осью. Для того чтобы совершить несколько поворотов в локальной системе осей, параллельных осям глобальной системы координат, надо совместить начала локальной и глобальной систем. Конкретнее, повороты могут быть выполнены с помощью следующей процедуры:

- переместить локальную систему осей так, чтобы начала локальной и глобальной систем совпали;
- выполнить требуемые повороты;
- переместить локальную систему осей обратно в исходное положение. Ниже эта процедура иллюстрируется на примере.

Пример 3-9 Несколько относительных вращений

Рассмотрим снова параллелепипед, изображенный на рис. 3-5 а. Пусть его требуется повернуть сначала на угол $\phi = -45^\circ$ вокруг оси y' , а затем на угол $\theta = +30^\circ$ вокруг оси x' . Для этого надо совместить начала систем координат $x'y'z'$ и xyz , выполнить необходимые повороты, а результат вернуть обратно в исходное положение.

Объединенное преобразование имеет вид:

$$[X'] = [X] [T] = [X] [Tr] [R_y] [R_x] [Tr]^{-1}.$$

Конкретнее,

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -x_c & -y_c & -z_c & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_c & y_c & z_c & 1 \end{bmatrix},$$

где ϕ и θ обозначают углы поворотов вокруг осей y' и x' , соответственно.

Объединив эти матрицы, получим

$$[T] = \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & -\sin \phi \cos \theta & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ \sin \phi & -\cos \phi \sin \theta & \cos \phi \cos \theta & 0 \\ x_c(1 - \cos \phi) & -x_c \sin \phi \sin \theta & x_c \sin \phi \cos \theta & 1 \\ -z_c \sin \phi & +y_c(1 - \cos \theta) & -y_c \sin \theta & 0 \\ +z_c \cos \phi \sin \theta & +z_c(1 - \cos \phi \cos \theta) & +z_c \cos \phi \cos \theta & 0 \end{bmatrix}. \quad (3-17)$$

Тогда преобразованные координатные векторы равны

$$[X'] = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.707 & -0.354 & 0.612 & 0 \\ 0 & 0.866 & 0.5 & 0 \\ -0.707 & -0.354 & 0.612 & 0 \\ 1.5 & 1.262 & -1.087 & 1 \end{bmatrix},$$

$$[X'] = \begin{bmatrix} 0.793 & 1.067 & 1.25 & 1 \\ 1.5 & 0.713 & 1.862 & 1 \\ 1.5 & 1.579 & 2.362 & 1 \\ 0.793 & 1.933 & 1.75 & 1 \\ 1.5 & 1.421 & 0.638 & 1 \\ 2.207 & 1.067 & 1.25 & 1 \\ 2.207 & 1.933 & 1.75 & 1 \\ 1.5 & 2.287 & 1.138 & 1 \end{bmatrix}.$$

Результат изображен на рис. 3-6.

3-9 ПОВОРОТ ВОКРУГ ПРОИЗВОЛЬНОЙ ОСИ В ПРОСТРАНСТВЕ

Обобщенный случай поворота вокруг произвольной оси в пространстве встречается часто, например в робототехнике, мультиплексии, моделировании. Следуя логике предыдущего обсуждения, поворот вокруг произвольной оси в пространстве выполняется с помощью переноса и простых поворотов вокруг координатных осей. Так как метод поворота вокруг координатной оси известен, то основная идея заключается в том, чтобы совместить произвольную ось вращения с одной из координатных осей.

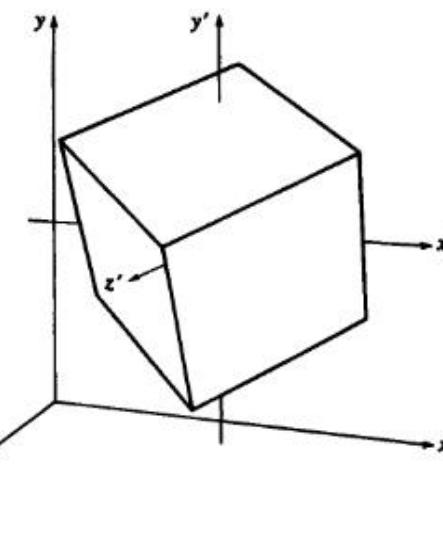


Рис. 3-6 Несколько поворотов относительно локальной системы координат.

Предположим, что произвольная ось в пространстве проходит через точку (x_0, y_0, z_0) с направляющим вектором (c_x, c_y, c_z) . Поворот вокруг этой оси на некоторый угол δ выполняется по следующему правилу:

- выполнить перенос так, чтобы точка (x_0, y_0, z_0) находилась в начале системы координат;
- выполнить соответствующие повороты так, чтобы ось вращения совпала с осью z^1 ;
- выполнить поворот на угол δ вокруг оси z ;
- выполнить преобразование, обратное тому, что позволило совместить ось вращения с осью z ;
- выполнить обратный перенос.

В общем случае для того, чтобы произвольная ось, проходящая через начало координат, совпала с одной из координатных осей, необходимо сделать два последовательных поворота вокруг двух других координатных осей. Для совмещения произвольной оси вращения с осью z сначала выполним поворот вокруг оси x , а затем вокруг оси y . Чтобы определить угол поворота α вокруг оси x , используемый для перевода произвольной оси в плоскость xz , спроектируем сначала на плоскость yz направляющий единичный вектор этой оси (рис. 3-7 a). Компоненты y и z спроектированного вектора равны c_y - и c_z -компонентам единичного направляющего вектора оси вращения.

Из рис. 3-7 a следует, что

$$d = \sqrt{c_y^2 + c_z^2} \quad (3-18)$$

¹ Выбор координатной оси (в данном случае z) произволен.

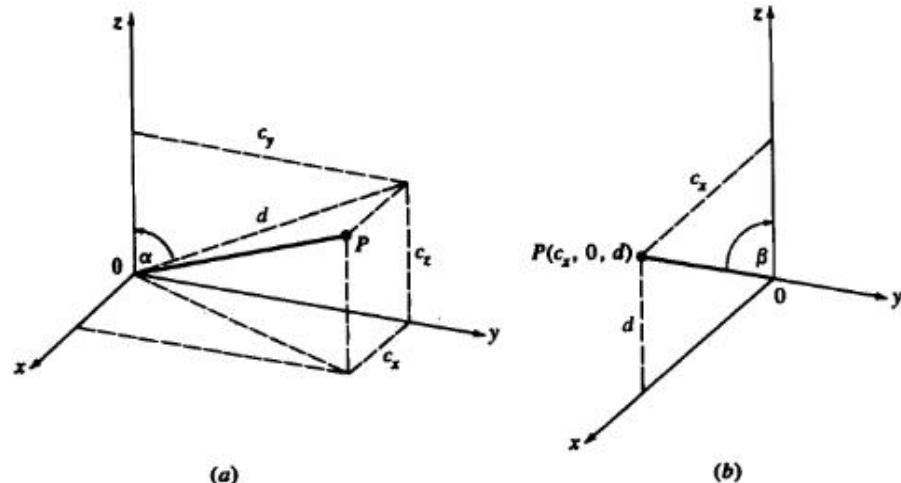


Рис. 3-7 Повороты, необходимые для совмещения с осью z единичного вектора OP .
(a) Поворот вокруг x ; (b) поворот вокруг y .

и

$$\cos \alpha = \frac{c_z}{d}, \quad \sin \alpha = \frac{c_y}{d}. \quad (3-19)$$

После перевода в плоскость zz с помощью поворота вокруг оси x , z -компоненты единичного вектора равна d , а x -компонента равна c_x , т.е. x -компоненте направляющего вектора, как это показано на рис. 3-7 b. Длина единичного вектора равна, конечно, 1. Таким образом, угол поворота β вокруг оси y , необходимый для совмещения произвольной оси с осью z , равен

$$\cos \beta = d, \quad \sin \beta = c_x. \quad (3-20)$$

Тогда полное преобразование можно представить в виде

$$[M] = [T][R_x][R_y][R_\delta][R_y]^{-1}[R_x]^{-1}[T]^{-1}, \quad (3-21)$$

где матрица переноса равна

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -x_0 & -y_0 & -z_0 & 1 \end{bmatrix}. \quad (3-22)$$

Матрица преобразования поворота вокруг оси x

$$[R_x] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c_x/d & c_y/d & 0 \\ 0 & -c_y/d & c_z/d & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3-23)$$

и вокруг оси y

$$[R_y] = \begin{bmatrix} \cos(-\beta) & 0 & -\sin(-\beta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(-\beta) & 0 & \cos(-\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} d & 0 & c_x & 0 \\ 0 & 1 & 0 & 0 \\ -c_x & 0 & d & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-24)$$

И, наконец, вращение вокруг произвольной оси задается матрицей поворота вокруг оси z

$$[R]_\delta = \begin{bmatrix} -\cos \delta & \sin \delta & 0 & 0 \\ -\sin \delta & \cos \delta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-25)$$

На практике углы α и β не вычисляются явным образом. Элементы матриц поворотов $[R_x]$ и $[R_y]$ в (3-21) получаются из уравнений (3-18)–(3-20) за счет выполнения двух операций деления и извлечения квадратного корня. Хотя данные результаты были разработаны для произвольной оси в первом квадранте, они применимы во всех квадрантах.

Если компоненты направляющего вектора произвольной оси неизвестны, то, зная вторую точку (x_1, y_1, z_1) на оси, их можно определить, нормализовав вектор, соединяющий первую и вторую точки. Более точно вектор оси из (x_0, y_0, z_0) в (x_1, y_1, z_1) равен

$$[V] = [(x_1 - x_0) (y_1 - y_0) (z_1 - z_0)].$$

Нормализация дает компоненты направляющего вектора:

$$[c_x \quad c_y \quad c_z] = \frac{[(x_1 - x_0) \quad (y_1 - y_0) \quad (z_1 - z_0)]}{[(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2]^{\frac{1}{2}}}. \quad (3-26)$$

Более полно этот метод иллюстрируется в следующем примере.

Пример 3-10 Поворот вокруг произвольной оси

Рассмотрим куб с одним отсеченным углом, изображенный на рис. 3-8 a. Координатные векторы вершин равны

$$[X] = \begin{bmatrix} 2 & 1 & 2 & 1 \\ 3 & 1 & 2 & 1 \\ 3 & 1.5 & 2 & 1 \\ 2.5 & 2 & 2 & 1 \\ 2 & 2 & 2 & 1 \\ 2 & 1 & 1 & 1 \\ 3 & 1 & 1 & 1 \\ 3 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 3 & 2 & 1.5 & 1 \end{bmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \\ I \\ J \end{matrix}$$

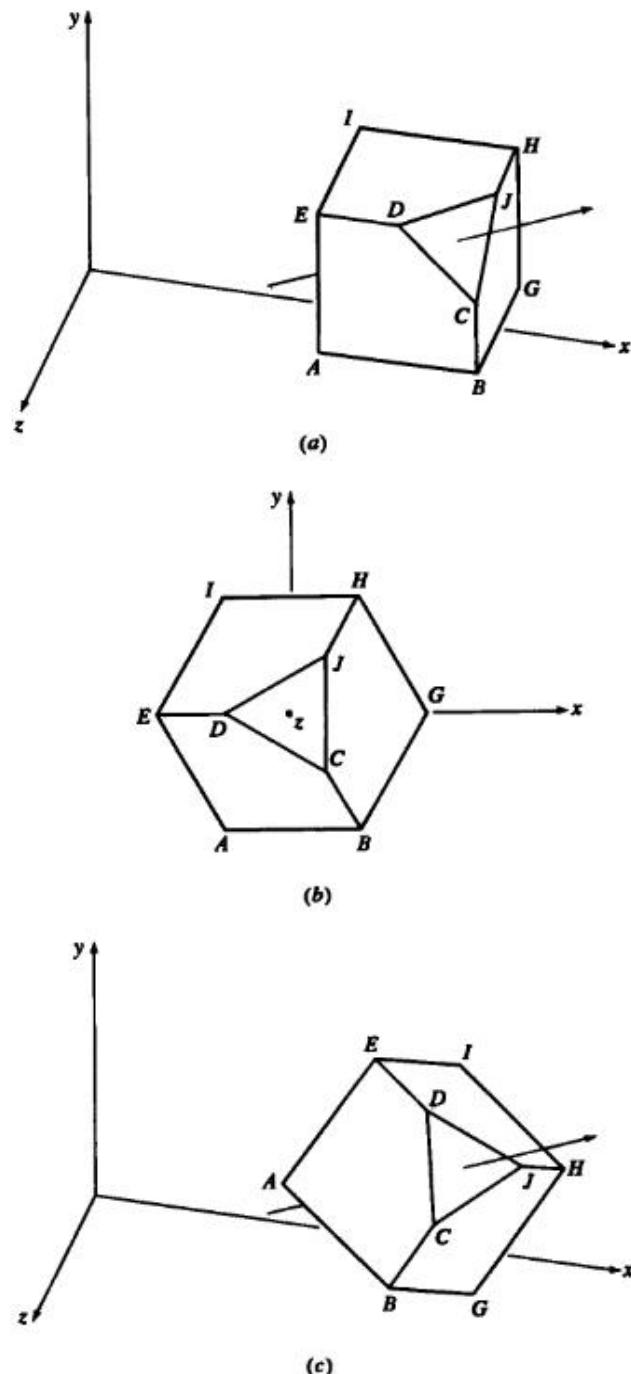


Рис. 3-8 Поворот вокруг произвольной оси.

Куб необходимо повернуть на -45° вокруг оси, проходящей через точку F и противоположный угол на диагонали. Ось направлена из точки F в противоположный угол и проходит через центр угловой грани.

Сначала определим компоненты направляющего единичного вектора оси вращения. Учитывая, что угол, отрезанный треугольником CDJ также лежит на оси, из (3-26) следует, что

$$\begin{bmatrix} c_x & c_y & c_z \end{bmatrix} = \frac{[(3-2) \quad (2-1) \quad (2-1)]}{((3-2)^2 + (2-1)^2 + (2-1)^2)^{1/2}} = \\ = [1/\sqrt{3} \quad 1/\sqrt{3} \quad 1/\sqrt{3}]$$

С помощью уравнений (3-18)–(3-20) получаем

$$d = \sqrt{(1/\sqrt{3})^2 + (1/\sqrt{3})^2} = \sqrt{2/3}$$

и

$$\alpha = \cos^{-1}(1/\sqrt{3}/\sqrt{2/3}) = \cos^{-1}(1/\sqrt{2}) = 45^\circ, \\ \beta = \cos^{-1}(\sqrt{2/3}) = 35, 26^\circ.$$

Так как точка F лежит на оси вращения, матрица переноса равна

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & -1 & -1 & 1 \end{bmatrix}.$$

Тогда матрицы поворотов, совмещающих произвольную ось вращения с осью z , имеют вид:

$$[R_z] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

и

$$[R_y] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ 0 & 1 & 0 & 0 \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Обратные матрицы $[R_z]^{-1}$, $[R_y]^{-1}$ и $[T]^{-1}$ получаются с помощью подстановки значений $-\alpha$, $-\beta$ и (x_0, y_0, z_0) для углов α , β и $(-x_0, -y_0, -z_0)$, соответственно, в уравнения (3-22)–(3-24).

Объединив $[T]$, $[R_x]$, $[R_y]$, получим

$$[M] = [T][R_x][R_y] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -2/\sqrt{6} & 0 & -4\sqrt{3} & 1 \end{bmatrix}.$$

Преобразованные промежуточные координаты равны

$$[X][M] = \begin{bmatrix} -0.408 & -0.707 & 0.577 & 1 \\ 0.408 & -0.707 & 1.155 & 1 \\ 0.204 & -0.354 & 1.443 & 1 \\ -0.408 & 0 & 1.443 & 1 \\ -0.816 & 0 & 0.155 & 1 \\ 0 & 0 & 0 & 1 \\ 0.816 & 0 & 0.577 & 1 \\ 0.408 & 0.707 & 1.155 & 1 \\ -0.408 & 0.707 & 0.577 & 1 \\ 0.204 & 0.354 & 1.443 & 1 \end{bmatrix}$$

Этот промежуточный результат изображен на рис. 3-8 б. Заметим, что точка F равна $(0, 0, 0)$.

Вращение вокруг произвольной оси теперь эквивалентно вращению вокруг оси z . Следовательно (см. (3-7)),

$$[R_\delta] = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразованный объект возвращается в его «исходное» местоположение в пространстве с помощью преобразования

$$[M]^{-1} = [R_y]^{-1}[R_x]^{-1}[T]^{-1} = \begin{bmatrix} 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 \\ 2 & 1 & 1 & 1 \end{bmatrix}.$$

Этот результат можно получить либо обращением и объединением матриц, составляющих $[M]$, либо обращением самой матрицы $[M]$. Заметим, между прочим, что $[R_x]$ и $[R_y]$ представляют только повороты. Верхняя левая (3×3) -подматрица $[M]^{-1}$ является точным транспонированием верхней левой (3×3) -подматрицы $[M]$.

Координаты результата равны

$$[X][M][R_\delta][M^{-1}] = \begin{bmatrix} 1.689 & 1.506 & 1.805 & 1 \\ 2.494 & 1.195 & 2.311 & 1 \\ 2.747 & 1.598 & 2.155 & 1 \\ 2.598 & 2.155 & 1.747 & 1 \\ 2.195 & 2.311 & 1.494 & 1 \\ 2 & 1 & 1 & 1 \\ 2.805 & 0.689 & 1.506 & 1 \\ 3.311 & 1.494 & 1.195 & 1 \\ 2.506 & 1.805 & 0.689 & 1 \\ 3.155 & 1.747 & 1.598 & 1 \end{bmatrix},$$

где

$$[M][R_\delta][M]^{-1} = \begin{bmatrix} 0.805 & -0.311 & 0.506 & 0 \\ 0.506 & 0.805 & -0.311 & 0 \\ -0.311 & 0.506 & 0.805 & 0 \\ 0.195 & 0.311 & -0.506 & 1 \end{bmatrix}.$$

На рис. 3-8 с изображен преобразованный объект.

3-10 ОТРАЖЕНИЕ ОТНОСИТЕЛЬНО ПРОИЗВОЛЬНОЙ ПЛОСКОСТИ

Преобразования, заданные в уравнениях (3-11)–(3-13), осуществляют отражение относительно координатных плоскостей $x = 0$, $y = 0$, $z = 0$ соответственно. Часто возникает необходимость отразить объект относительно произвольной плоскости. И снова это можно сделать с помощью процедуры, объединяющей ранее определенные простые преобразования. Один из возможных методов состоит в следующем:

- перенести точку P , принадлежащую плоскости отражения, в начало координат;
- повернуть вектор нормали к плоскости отражения в начале координат до совпадения с осью $+z$ (см. разд. 3-9, уравнения (3-23) и (3-24)), теперь плоскость отражения будет совпадать с координатной плоскостью $z = 0$;
- применяя уже известные преобразования, отразить объект относительно координатной плоскости $z = 0$ (см. (3-11));
- чтобы получить результаты, необходимо выполнить преобразования, обратные к описанным в первых двух пунктах.

Тогда общее преобразование описывается матрицей

$$[M] = [T][R_x][R_y][Rfl_{tz}][R_y]^{-1}[R_x]^{-1}[T]^{-1},$$

где матрицы $[T]$, $[R_x]$, $[R_y]$ задаются уравнениями (3-22)–(3-24) соответственно, Rfl_{tz} – матрица отражения относительно плоскости $z = 0$, $(x_0, y_0, z_0) = (P_z, P_y, P_z)$ – координаты точки P на плоскости отражения; а (c_x, c_y, c_z) есть вектор нормали к плоскости отражения¹.

Пример более подробно проиллюстрирует данный метод.

Пример 3-11 Отражение

Снова рассмотрим куб с одним отсеченным углом, как это показано на рис. 3-8 а. Отразим куб относительно плоскости, содержащей треугольник CDJ .

Выбрав точку C для перемещения в начало координат и вспомнив координатные векторы для куба, получим матрицу переноса

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & -3/2 & -2 & 1 \end{bmatrix}.$$

¹ Если известно уравнение плоскости отражения, $ax + by + cz + d = 0$, то единичный вектор нормали к плоскости равен

$$[\hat{n}] = [c_x \quad c_y \quad c_z] = \frac{[a \quad b \quad c]}{\sqrt{a^2 + b^2 + c^2}}.$$

Подробности можно найти в [3-1].

Нормаль к плоскости отражения получим, используя координатные векторы C, D, J (см. [3-1]). Конкретнее, взяв векторное произведение векторов CJ и CD до переноса, получим

$$\begin{aligned} n &= ([J] - [C]) \times ([D] - [C]) = \\ &= [(3 - 3) \quad (2 - 1.5) \quad (1.5 - 2)] \times [(2.5 - 3) \quad (2 - 1.5) \quad (2 - 2)] = \\ &= [0 \quad 1/2 \quad -1/2] \times [-1/2 \quad 1/2 \quad 0] = [1/4 \quad 1/4 \quad 1/4]. \end{aligned}$$

Нормализация дает

$$\hat{n} = [1/\sqrt{3} \quad 1/\sqrt{3} \quad 1/\sqrt{3}].$$

Используя уравнения (3-19) и (3-20), получим

$$d = \sqrt{n_y^2 + n_z^2} = \sqrt{(1/\sqrt{3})^2 + (1/\sqrt{3})^2} = \sqrt{2/3}$$

и $\alpha = 45^\circ$, $\beta = 35.26^\circ$. Матрицы поворотов для совмещения нормали в точке C с осью z имеют вид (см. (3-23) и (3-24))

$$[R_x] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad [R_y] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ 0 & 1 & 0 & 0 \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Матрицы $[R_x]^{-1}$, $[R_y]^{-1}$ и $[T]^{-1}$ можно получить подстановкой $-\alpha$, $-\beta$ и $[x_0 \quad y_0 \quad z_0] = [C]$ в (3-22)–(3-24).

Объединение $[T]$, $[R_x]$ и $[R_y]$ дает

$$[M] = [T][R_x][R_y] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -5/2\sqrt{6} & 1/2\sqrt{2} & -13/2\sqrt{3} & 1 \end{bmatrix}.$$

Преобразованные промежуточные точки равны

$$[X][M] = \begin{bmatrix} -0.612 & -0.354 & -0.876 & 1 \\ 0.204 & -0.354 & -0.287 & 1 \\ 0 & 0 & 0 & 1 \\ -0.612 & 0.354 & 0 & 1 \\ -1.021 & 0.354 & -0.287 & 1 \\ -0.204 & 0.354 & -1.443 & 1 \\ 0.612 & 0.354 & -0.876 & 1 \\ 0.204 & 1.061 & -0.287 & 1 \\ -0.612 & 1.061 & -0.876 & 1 \\ 0 & 0.707 & 0 & 1 \end{bmatrix}.$$

Этот промежуточный результат изображен на рис. 3-9 b. Заметим, что точка C находится в начале системы координат, а ось z направлена на нас.

Отражение относительно произвольной плоскости теперь эквивалентно отражению относительно плоскости $z = 0$. Следовательно (см. (3-11)),

$$[R_{flit}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

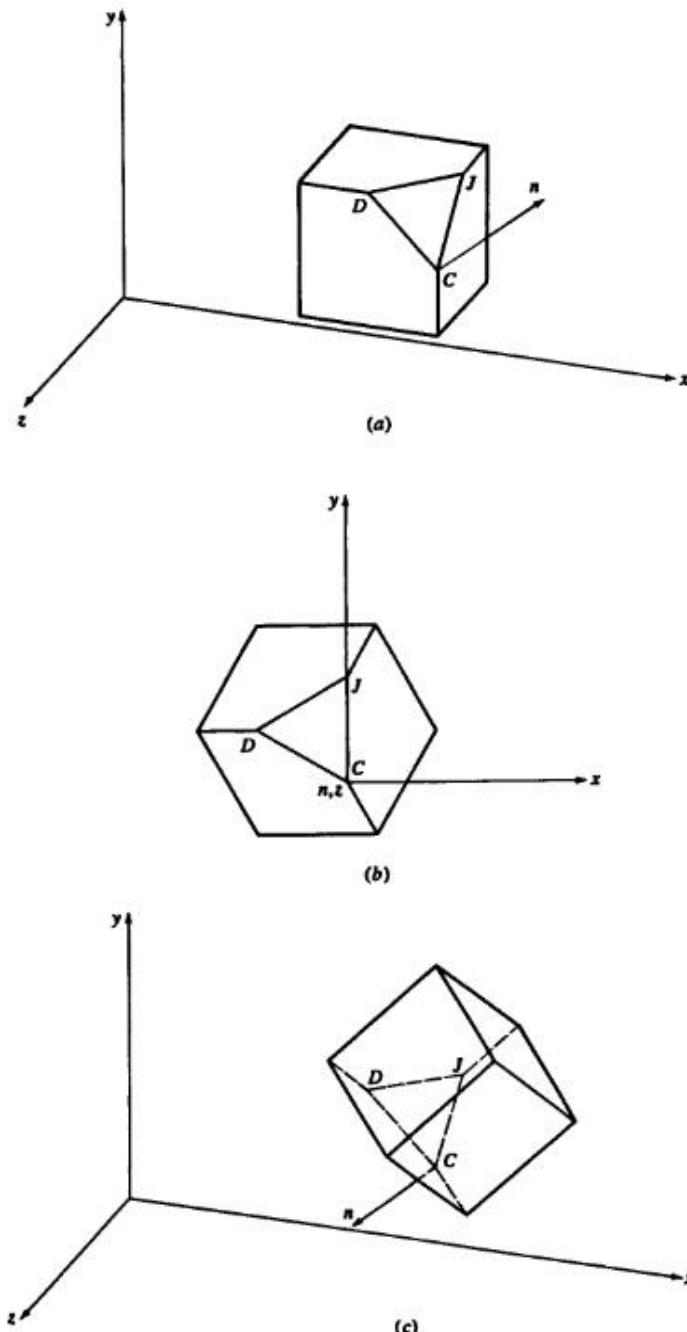


Рис. 3-9 Отражение относительно произвольной плоскости.

Для возврата преобразованного объекта в «исходное» положение в пространстве требуется преобразование

$$[M]^{-1} = [R_y]^{-1}[R_x]^{-1}[T]^{-1} = \begin{bmatrix} 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 \\ 3 & 3/2 & 2 & 1 \end{bmatrix}.$$

Результатирующие координатные векторы таковы:

$$[X][M][RfIt][M]^{-1} = \begin{bmatrix} 3 & 2 & 3 & 1 \\ 10/3 & 4/3 & 7/3 & 1 \\ 3 & 3/2 & 2 & 1 \\ 5/2 & 2 & 2 & 1 \\ 7/3 & 7/3 & 7/3 & 1 \\ 11/3 & 8/3 & 8/3 & 1 \\ 4 & 2 & 2 & 1 \\ 10/3 & 7/3 & 4/3 & 1 \\ 3 & 3 & 2 & 1 \\ 3 & 2 & 3/2 & 1 \end{bmatrix},$$

где

$$[M][RfIt][M]^{-1} = \begin{bmatrix} 1/3 & -2/3 & -2/3 & 0 \\ -2/3 & 1/3 & -2/3 & 0 \\ -2/3 & -2/3 & 1/3 & 0 \\ 13/3 & 13/3 & 13/3 & 1 \end{bmatrix}.$$

На рис. 3-9 с показан преобразованный объект.

Как показано в этом и предыдущих разделах, сложные преобразования можно легко построить с помощью простых базовых преобразований. Такой подход даже предпочтительнее, так как он уменьшает вероятность появления ошибок и более эффективен с вычислительной точки зрения, нежели прямой математический подход.

3-11 АФФИННАЯ И ПЕРСПЕКТИВНАЯ (НАЧЕРТАТЕЛЬНАЯ) ГЕОМЕТРИЯ

К настоящему времени разработана математическая теория как для перспективной (начертательной), так и для аффинной геометрии. Теоремы аффинной геометрии идентичны теоремам геометрии Евклида. И в той и в другой науках важным понятием является параллелизм. В перспективной геометрии прямые в общем случае не параллельны.

Аффинное преобразование есть комбинация линейных преобразований, например поворота и последующего переноса. Для аффинного преобразования последний столбец в обобщенной 4×4 -матрице равен $[0 \ 0 \ 0 \ 1]^T$. В противном случае, как это показано далее в разделе 3-15, преобразованная однородная координата h не равна единице и нет взаимно однозначного соответствия между аффинным преобразованием и (4×4) -матричным оператором. Аффинные преобразования образуют полезное подмножество билинейных преобразований, так



Рис. 3-10 Иерархия плоских геометрических проекций.

как произведение двух аффинных преобразований также аффинно. Это свойство позволяет скомбинировать общее преобразование множества точек относительно произвольной системы координат при сохранении значения единицы для однородной координаты h .

Так как евклидова геометрия изучается в школах многие годы, то методы рисования и черчения, основывающиеся на евклидовой геометрии, стали стандартными методами графического сообщения. Хотя художниками и архитекторами часто используются перспективные виды для создания более реалистического изображения, в технической работе они используются редко из-за трудности их ручного конструирования. Однако при использовании для задания объекта однородных координат, аффинные и перспективные преобразования вычисляются одинаково легко.

И аффинные, и перспективные преобразования трехмерны, т. е. являются преобразованиями одного трехмерного пространства в другое. Однако для наблюдения результатов на двумерной поверхности требуется проецирование из трехмерного пространства в двумерное. Результат этого проецирования называется плоской геометрической проекцией. На рис. 3-10 изображена иерархия плоских геометрических проекций. Матрица проецирования из трехмерного пространства в двумерное всегда содержит столбец из нулей, следовательно, детерминант этого преобразования всегда равен нулю.

Плоские геометрические проекции объектов образуются пересечением прямых, называемых проекторами, с плоскостью, называемой плоскостью проекции. Проекторы — это прямые, проходящие через произвольную точку, называемую центром проекции, и каждую точку объекта. Если центр проекции расположен в конечной точке трехмерного пространства, получается перспективная проекция. Если центр расположен в бесконечности, то все проекторы параллельны и результат является параллельной проекцией. Плоские геометрические проекции пред-

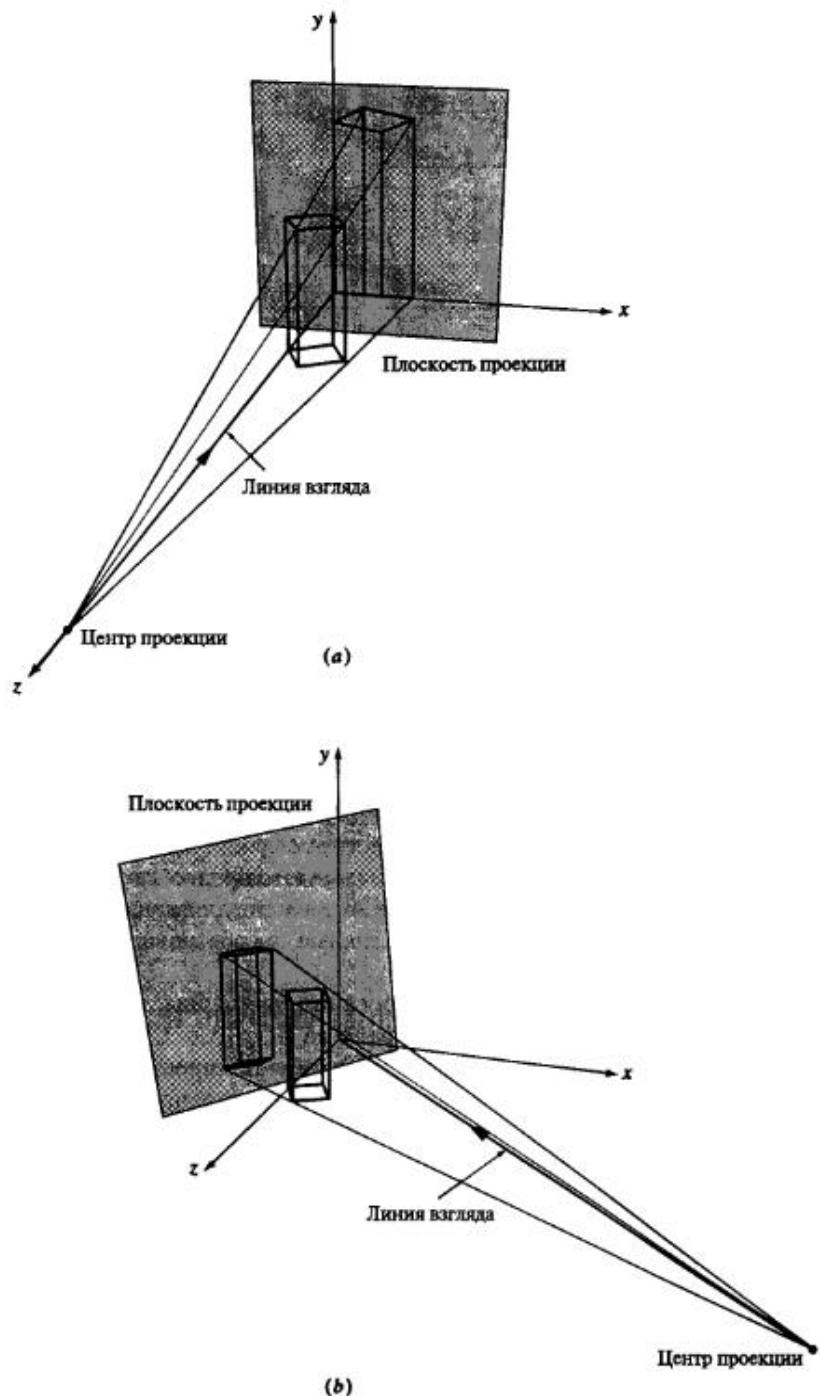


Рис. 3-11 Плоские проекции. (a) Фиксирован центр проекции; (b) фиксирован объект.

ставляют основу описательной геометрии. Неплоские и негеометрические проекции также полезны; они широко используются в картографии.

Для разработки различных преобразований, представленных на рис. 3-10, можно использовать два разных подхода. В первом предполагается, что центр проекции или точка зрения фиксирована, а плоскость проекции перпендикулярна каждому проектору, как это показано на рис. 3-11 a. Для получения требуемого вида манипулируют объектом. Во втором подходе предполагается, что объект фиксирован, центр проекции может как угодно перемещаться в трехмерном пространстве, а плоскость проекции не обязательно перпендикулярна направлению взгляда. На рис. 3-11 b приведен пример этого. Оба подхода математически эквивалентны.

Первый подход напоминает действия наблюдателя, которого попросили описать небольшой объект, например, книгу. Он берет объект в руки, поворачивает и перемещает его для того, чтобы изучить все его стороны. В этом случае центр проекции фиксирован и манипулируют объектом. Второй подход напоминает действия наблюдателя, которого попросили описать большой объект, например автомобиль. Наблюдатель ходит вокруг объекта, чтобы осмотреть его с разных сторон, взбирается на лестницу для осмотра верха и опускается на колени для осмотра его нижней части. В этом случае объект фиксирован, а центр проекции и точка зрения перемещаются.

В процессе конструирования или изображения объекта на графическом дисплее компьютера местоположение глаза обычно фиксировано, а плоскость проекции, т. е. поверхность, обычно перпендикулярна направлению взгляда. Следовательно, в данном случае более подходит первый подход. Тем не менее, если графический дисплей используется для представления движения какого-нибудь транспортного средства или наблюдатель движется в генерированной компьютером модели, как в случае использования тренажера, или если наблюдатель прогуливается в архитектурной модели, тогда больше подходит второй подход.

В этой книге используются фиксированный центр проекции и движущийся объект. Метод с фиксированным объектом и движущимся центром проекции хорошо проработан Калбомом и Пейсиореком ([3-2]).

Мы начнем наше обсуждение плоских геометрических проекций (см. рис. 3-10) с рассмотрения параллельных проекций.

3-12 ОРТОГРАФИЧЕСКИЕ ПРОЕКЦИИ

Самой простой из параллельных проекций является ортографическая проекция, используемая обычно в инженерных чертежах. В этом случае точно изображаются правильные или «истинные» размер и форма одной плоской грани объекта. Ортографические проекции — это проекции на одну из координатных плоскостей $z = 0$, $y = 0$ или $x = 0$. Матрица проекции на плоскость $z = 0$ имеет вид

$$[P_z] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-27)$$

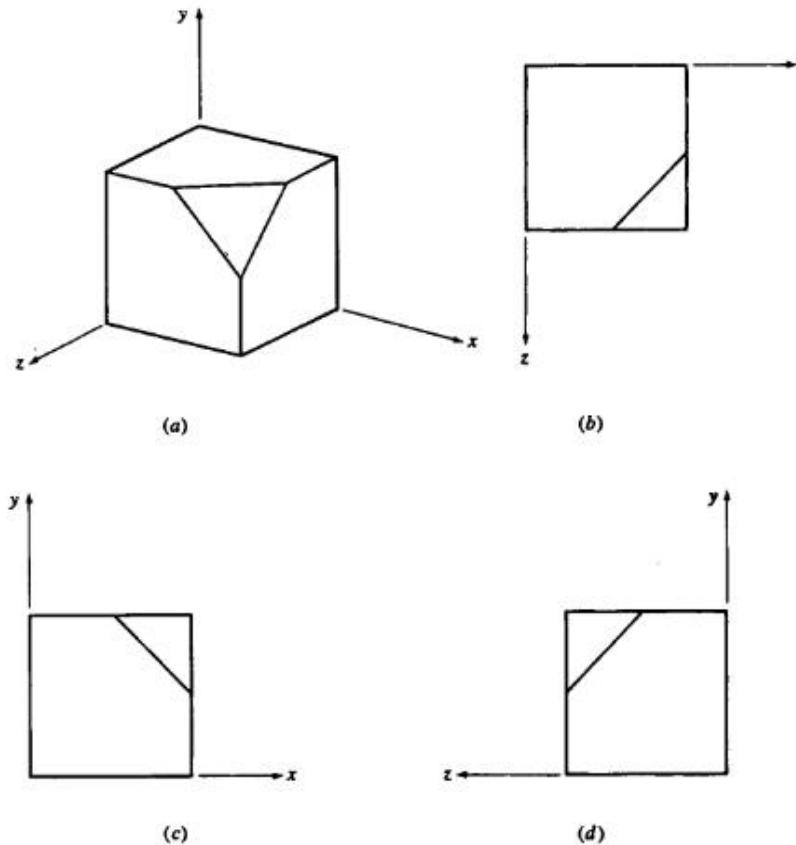


Рис. 3-12 Ортографические проекции объекта на рис. 3-12а на плоскости: (b) $y = 0$, (c) $z = 0$ и (d) $x = 0$.

Заметим, что в третьем столбце (столбце z) все элементы нулевые. Следовательно, в результате преобразования z -координата координатного вектора станет равной нулю.

Аналогичным образом матрицы для проекций на плоскости $x = 0$ и $y = 0$ равны

$$[P_x] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3-28)$$

$$[P_y] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3-29)$$

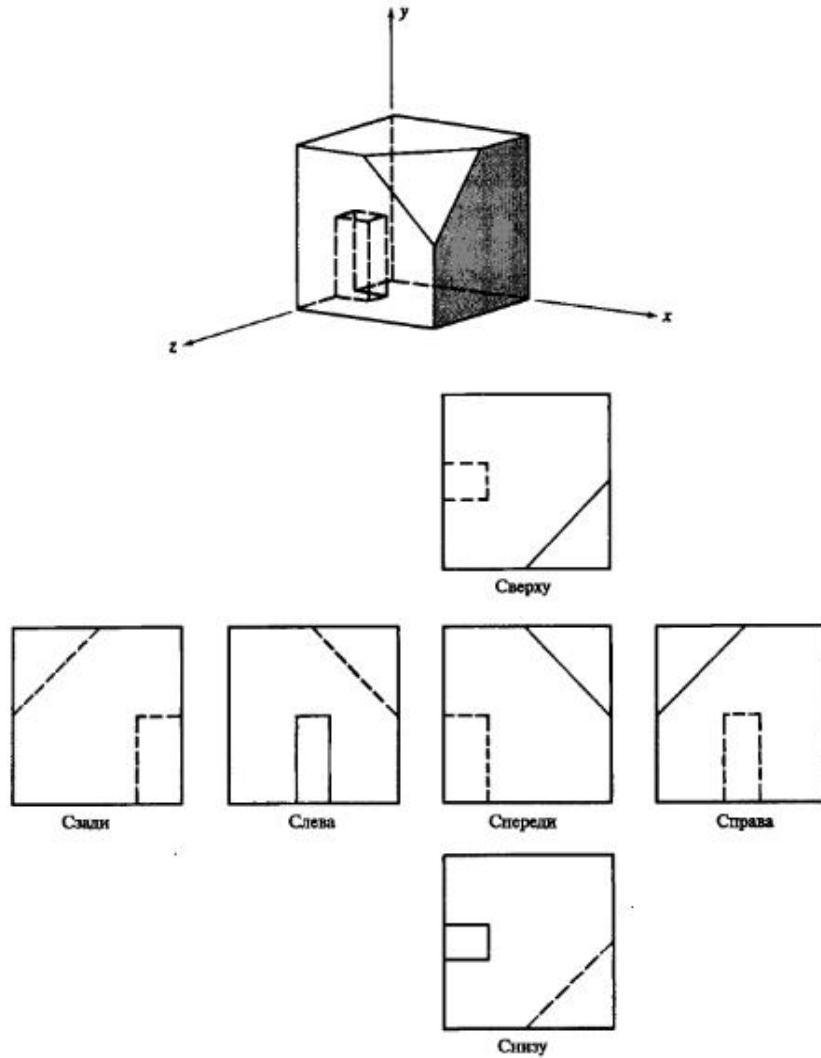


Рис. 3-13 Многовидовая ортографическая проекция.

Ортографические проекции объекта на рис. 3-12 а на плоскости $x = 0$, $y = 0$ и $z = 0$ из центров проекций, расположенных в бесконечности на $+x$ -, $+y$ - и $+z$ -осях соответственно, изображены на рис. 3-12 b, 3-12 c и 3-12 d соответственно.

Одна ортографическая проекция не дает достаточной информации для визуального и практического воссоздания формы объекта. Необходимо, следовательно, наличие нескольких ортографических проекций. Эти многовидовые ортографические проекции обычно¹ организуются так, как это показано на рис. 3-13.

¹ Это соглашение относится к Соединенным Штатам.

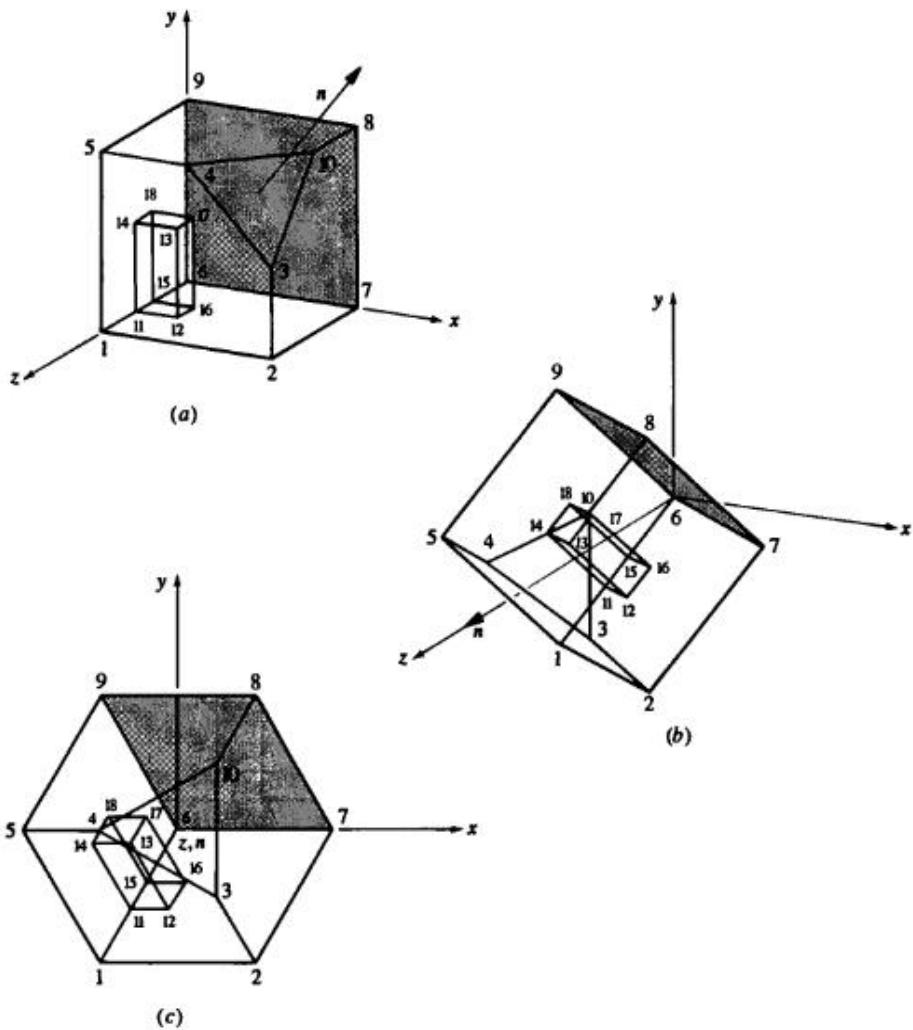


Рис. 3-14 Построение вспомогательного вида. (а) Тринетрический вид; (б) нормаль совмещена с осью z ; (с) спроецировано на плоскость $z = 0$.

Виды спереди, справа и сверху получаются проецированием на плоскости $z = 0$, $x = 0$ и $y = 0$ из центров проекции, расположенных в бесконечности на $+z$, $+x$ и $+y$ осях. Виды сзади, слева и снизу получаются проецированием на плоскости $z = 0$, $x = 0$, $y = 0$ из центров проекции, расположенных в бесконечности на $-z$, $-x$ и $-y$ осях. Координатные оси обычно не изображаются на этих видах.

Как показано на рис. 3.13, невидимые линии принято изображать пунктиром. Для адекватного представления формы объекта обычно не требуются все шесть видов. Наиболее часто используются виды спереди, сверху и справа. Даже если используются не все шесть видов, оставшиеся располагаются в указанных местах.

Виды спереди и с боков иногда называют главным и боковыми фасадами, вид сверху иногда называют планом.

Интересно и важно отметить, что все шесть видов могут быть получены комбинациями отражения, вращения и переноса с последующим проецированием на плоскость $z = 0$ из центра проекции, расположенного в бесконечности на оси z . Например, вид сзади получается отражением относительно плоскости $z = 0$ и проецированием на эту плоскость. Аналогичным образом вид слева получается вращением вокруг оси y на угол $+90^\circ$ и проецированием на плоскость $z = 0$.

Для объектов с гранями, не параллельными одной из координатных плоскостей, стандартные ортографические виды не показывают правильную или истинную форму этих граней. Для этих целей используются вспомогательные виды. Вспомогательный вид образуется с помощью вращения и перемещения объекта так, чтобы нормаль к грани совпала с одной из координатных осей (см. разд. 3-9). Результат затем проецируется на координатную плоскость, перпендикулярную этой оси. На рис. 3-14 с изображена вспомогательная плоскость, показывающая истинную форму треугольного угла объекта на рис. 3.13.

Пример более полно объясняет эти построения.

Пример 3-12 Вспомогательный вид

Построим вспомогательный вид, изображающий правильную форму треугольного угла объекта на рис. 3-14 а. Координатные векторы объекта равны

$$[X] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0.5 & 1 & 1 \\ 0.5 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0.5 & 1 \\ 0 & 0 & 0.6 & 1 \\ 0.25 & 0 & 0.6 & 1 \\ 0.25 & 0.5 & 0.6 & 1 \\ 0 & 0.5 & 0.6 & 1 \\ 0 & 0 & 0.4 & 1 \\ 0.25 & 0 & 0.4 & 1 \\ 0.25 & 0.5 & 0.4 & 1 \\ 0 & 0.5 & 0.4 & 1 \end{bmatrix}.$$

Номера вершин, изображенные на рис. 3-14, соответствуют номерам строк матрицы $[X]$ координатных векторов.

Внешняя нормаль к треугольной грани имеет следующие направляющие косинусы

$$[c_x \ c_y \ c_z] = [1/\sqrt{3} \ 1/\sqrt{3} \ 1/\sqrt{3}]$$

и проходит через начало координат и точку [0.83333 0.83333 0.83333]. Вспомнив результаты разд. 3-9 и примера 3-10, получим, что нормаль совмещается с осью z с помощью поворота вокруг оси x на угол

$$\alpha = \arccos(c_z/d) = \arccos(1/\sqrt{2}) = +45^\circ$$

и последующего поворота вокруг оси y на угол

$$\beta = \arccos(d) = \arccos(2/\sqrt{6}) = +35,26^\circ.$$

Теперь общая матрица преобразования равна

$$[T] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Преобразованные координатные векторы равны

$$[X'] = [X][T] = \begin{bmatrix} -0.408 & -0.707 & 0.577 & 1 \\ 0.408 & -0.707 & 1.155 & 1 \\ 0.204 & -0.354 & 1.443 & 1 \\ -0.408 & 0 & 1.443 & 1 \\ -0.816 & 0 & 1.155 & 1 \\ 0 & 0 & 0 & 1 \\ 0.816 & 0 & 0.577 & 1 \\ 0.408 & 0.707 & 0.155 & 1 \\ -0.408 & 0.707 & 0.577 & 1 \\ 0.204 & 0.354 & 1.443 & 1 \\ -0.245 & -0.424 & 0.354 & 1 \\ -0.041 & -0.424 & 0.491 & 1 \\ -0.245 & -0.071 & 0.779 & 1 \\ -0.449 & -0.071 & 0.635 & 1 \\ -0.163 & -0.283 & 0.231 & 1 \\ 0.041 & -0.283 & 0.375 & 1 \\ -0.163 & 0.071 & 0.664 & 1 \\ -0.367 & 0.071 & 0.52 & 1 \end{bmatrix}$$

Результат изображен на рис. 3-14 б. Вспомогательный вид создается проецированием этого промежуточного результата на плоскость $z = 0$ с помощью (3-27), т. е.

$$[P_z] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Объединение матриц $[T]$ и $[P_z]$ дает

$$[T'] = [T][P_z] = \begin{bmatrix} 2/\sqrt{6} & 0 & 0 & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{3} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Обратите внимание на столбец из нулей. Вспомогательный вид создается затем с помощью

$$[X''] = [X][T'].$$

Матрица $[X'']$ та же самая, что и $[X']$ за исключением нулевого третьего столбца, т. е. действие проецирования состоит в отбрасывании координаты z . Результат показан на рис. 3-14 с. Невидимые линии изображены сплошными. Заметим, что показана правильная форма грани объекта — равностороннего треугольника.

Часто бывает необходимо для сложных объектов показать внутренние детали. Это выполняется с помощью сечений. Сечение образуется проведением через объект так называемой «секущей» плоскости, удалением части объекта по одну сторону от этой плоскости и проецированием оставшейся части объекта на плоскость сечения. Как и ранее, этот вид можно сделать с помощью совмещения с одной из координатных осей нормали к плоскости сечения (см. разд. 3-9), отсечения объекта с одной стороны от плоскости сечения (см. [3-1]) и, наконец, проецирования результата на координатную плоскость, перпендикулярную этой оси.

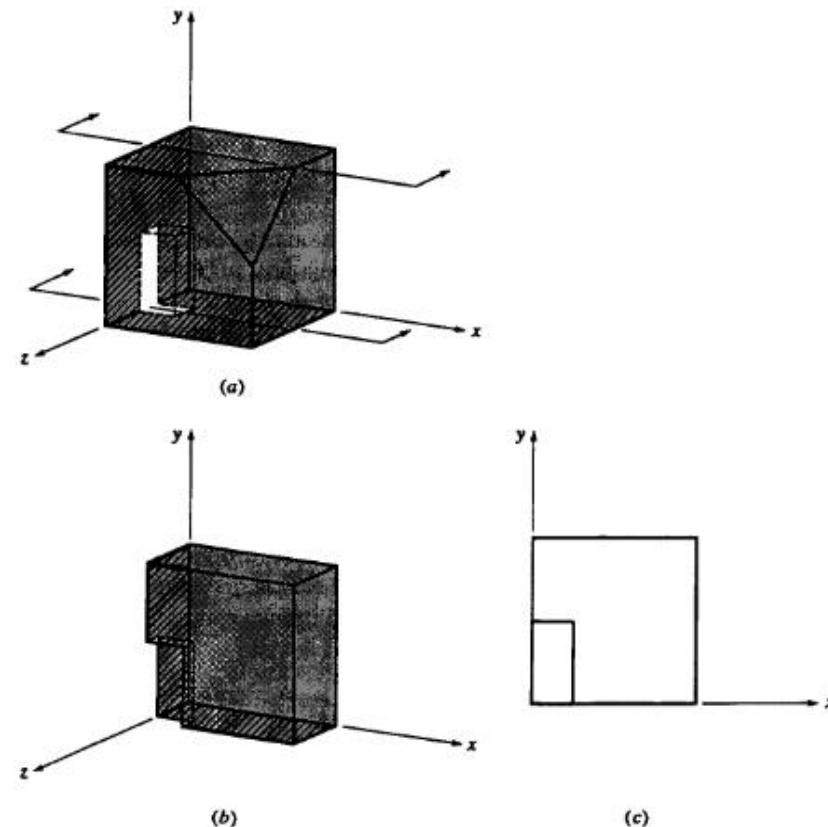


Рис. 3-15 Построение сечения. (а) Весь объект; (б) удалена часть между плоскостью сечения и центром проецирования; (с) проекция на плоскость $z = 0$.

На рис. 3-15 изображена плоскость сечения, проходящая через выемку на левой стороне объекта с рис. 3-13. Стрелки используются для показа плоскости сечения и направления взгляда.

3-13 АКСОНОМЕТРИЧЕСКИЕ ПРОЕКЦИИ

Одна ортографическая проекция не может дать представления об общей трехмерной форме объекта. Это ограничение можно преодолеть с помощью аксонометрических проекций. Аксонометрическая проекция образуется манипулированием объекта с помощью поворотов и перемещений таким образом, что бы были видны по крайней мере три соседние грани¹. Результат затем проецируется с центром проекции, расположенным в бесконечности, на одну из координатных плоскостей, обычно на плоскость $z = 0$. Если грань не параллельна плоскости проекции, то аксонометрическая проекция не показывает истинную форму этой грани. Однако остаются постоянными относительные длины параллельных в исходном пространстве линий, т. е. параллельные линии одинаково укорачиваются (искажаются). Коэффициент искажения есть отношение длины проекции отрезка к его истинной длине. Представляют интерес три аксонометрические проекции: триметрическая, диметрическая и изометрическая, как это показано на рис. 3-10. В триметрической проекции меньше всего ограничений, а в изометрической — больше всего. В самом деле, как будет показано ниже, изометрическая проекция есть частный случай диметрической, а диметрическая проекция есть частный случай триметрической.

Триметрическая проекция строится произвольными поворотами вокруг любых координатных осей, совершаемыми в произвольном порядке, с последующим проецированием на плоскость $z = 0$. Большинство иллюстраций в этой книге представлены триметрическими проекциями. На рис. 3-16 изображено несколько таких проекций. Каждая из них построена с помощью поворота вокруг оси y , затем вокруг оси x и, наконец, параллельного проецирования на плоскость $z = 0$.

В общем случае для триметрической проекции коэффициенты искажения по каждой из проецируемых главных осей (x , y и z) не равны друг другу. Здесь термин главная ось используется в том смысле, что ось (или ребро) объекта в исходном пространстве параллельна одной из координатных осей x , y или z . Наложение ограничений на коэффициенты уменьшает диапазон триметрических проекций. Однако для любой конкретной триметрической проекции коэффициенты искажения вычисляют с помощью применения общей матрицы преобразования к единичным векторам вдоль главных осей. В частности,

$$[U][T] = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} [T] = \begin{bmatrix} x_z^* & y_z^* & 0 & 1 \\ x_y^* & y_y^* & 0 & 1 \\ x_z^* & y_z^* & 0 & 1 \end{bmatrix}, \quad (3-30)$$

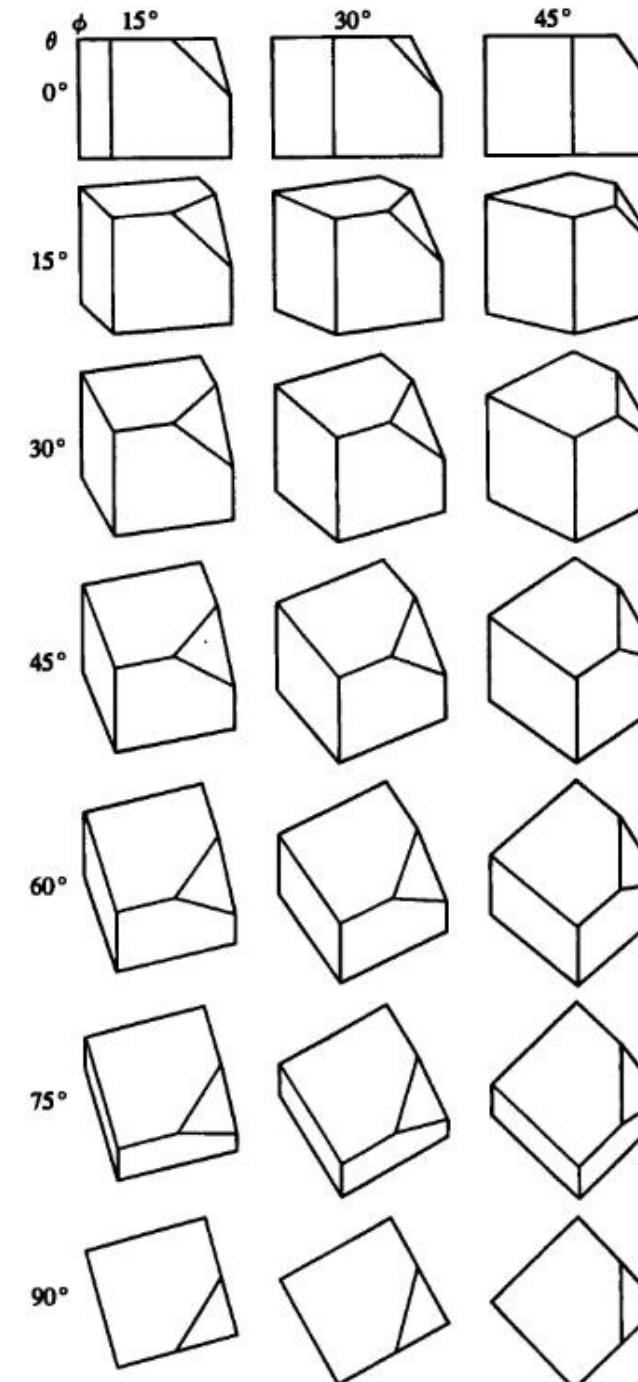


Рис. 3-16 Триметрические проекции.

¹ Минимальное количество граней, встречающихся для простых кубовидных объектов, подобных используемым в большинстве иллюстраций этой главы.

где $[U]$ есть матрица единичных векторов вдоль нетрансформированных осей x , y и z соответственно, а $[T]$ — общая матрица триметрической проекции. Тогда коэффициенты искажения вдоль спроектированных главных осей равны

$$f_x = \sqrt{x_x^{*2} + y_z^{*2}}, \quad (3-31a)$$

$$f_y = \sqrt{x_y^{*2} + y_y^{*2}}, \quad (3-31b)$$

$$f_z = \sqrt{x_z^{*2} + y_z^{*2}}. \quad (3-31c)$$

В примере 3-13 триметрическая проекция рассмотрена более подробно.

Пример 3-13 Триметрическая проекция

Рассмотрим центральную иллюстрацию рис. 3-16, построенную с помощью поворота на угол $\phi = 30^\circ$ вокруг оси y , затем поворота на угол $\theta = 45^\circ$ вокруг оси x и последующего параллельного проецирования на плоскость $z = 0$. Координатный вектор куба с отсеченым углом

$$[X] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0.5 & 1 & 1 \\ 0.5 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0.5 & 1 \end{bmatrix}.$$

Общая матрица триметрической проекции равна (см. (3-8), (3-6) и (3-27)):

$$\begin{aligned} [T] &= [R_y][R_x][P_z] = \\ &= \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & 0 \\ 0 & \cos \theta & 0 & 0 \\ \sin \phi & -\cos \phi \sin \theta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & \sqrt{2}/4 & 0 & 0 \\ 0 & \sqrt{2}/2 & 0 & 0 \\ 1/2 & -\sqrt{6}/4 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Таким образом, преобразованные координатные векторы имеют вид:

$$[X^*] = [X][T] = \begin{bmatrix} 0.5 & -0.612 & 0 & 1 \\ 1.366 & -0.259 & 0 & 1 \\ 1.366 & 0.095 & 0 & 1 \\ 0.933 & 0.272 & 0 & 1 \\ 0.5 & 0.095 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.866 & 0.354 & 0 & 1 \\ 0.866 & 1.061 & 0 & 1 \\ 0 & 0.707 & 0 & 1 \\ 1.116 & 0.754 & 0 & 1 \end{bmatrix},$$

3-13 Аксонометрические проекции

а коэффициенты искажения —

$$\begin{aligned} [U][T] &= \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3}/2 & \sqrt{2}/4 & 0 & 0 \\ 0 & \sqrt{2}/2 & 0 & 0 \\ 1/2 & -\sqrt{6}/4 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \sqrt{3}/2 & \sqrt{2}/4 & 0 & 1 \\ 0 & \sqrt{2}/2 & 0 & 1 \\ 1/2 & -\sqrt{6}/4 & 0 & 1 \end{bmatrix} \end{aligned}$$

и

$$\begin{aligned} f_x &= \sqrt{(\sqrt{3}/2)^2 + (\sqrt{2}/4)^2} = 0.935, \\ f_y &= \sqrt{2}/2 = 0.707, \\ f_z &= \sqrt{(1/2)^2 + (-\sqrt{6}/4)^2} = 0.791. \end{aligned}$$

Диметрическая проекция — это триметрическая проекция с двумя одинаковыми коэффициентами искажения, третий коэффициент может иметь любое значение. Диметрическая проекция строится с помощью поворота на угол ϕ вокруг оси y , затем поворота на угол θ вокруг оси x и проецирования на плоскость $z = 0$ с центром проекции, расположенным в бесконечности. Точные значения углов поворотов еще не известны. Воспользовавшись (3-8), (3-6) и (3-27), получим результатирующее преобразование:

$$\begin{aligned} [T] &= [R_y][R_x][R_z] \\ &= \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Объединение матриц дает

$$[T] = \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & 0 \\ 0 & \cos \theta & 0 & 0 \\ \sin \phi & -\cos \phi \sin \theta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-32)$$

Единичные векторы на главных осях x , y и z преобразуются в

$$[U^*] = [U][T] = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & 0 \\ 0 & \cos \theta & 0 & 0 \\ \sin \phi & -\cos \phi \sin \theta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$[U^*] = \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & 1 \\ 0 & \cos \theta & 0 & 1 \\ \sin \phi & -\cos \phi \sin \theta & 0 & 1 \end{bmatrix}. \quad (3-33)$$

Теперь квадрат длины подвергнутого преобразованию единичного вектора вдоль оси x , т. е. квадрат коэффициента искажения, равен

$$f_x^2 = (x_x^{*2} + y_x^{*2}) = \cos^2 \phi + \sin^2 \phi \sin^2 \theta. \quad (3-34)$$

Аналогично квадраты коэффициентов искажения по осям y и z равны

$$f_y^2 = (x_y^{*2} + y_y^{*2}) = \cos^2 \theta, \quad (3-35)$$

$$f_z^2 = (x_z^{*2} + y_z^{*2}) = \sin^2 \phi + \cos^2 \phi \sin^2 \theta. \quad (3-36)$$

Приравнивание коэффициентов искажения по осям¹ x и y дает одно уравнение с двумя неизвестными углами поворота ϕ и θ . А именно

$$\cos^2 \phi + \sin^2 \phi \sin^2 \theta = \cos^2 \theta.$$

Используя равенства $\cos^2 \phi = 1 - \sin^2 \phi$ и $\cos^2 \theta = 1 - \sin^2 \theta$, получаем

$$\sin^2 \phi = \sin^2 \theta / (1 - \sin^2 \theta). \quad (3-37)$$

Второе соотношение между ϕ и θ получим, фиксируя коэффициент искажения f_z вдоль оси z . Объединив уравнения (3-36) и (3-37) и воспользовавшись равенством $\cos^2 \phi = 1 - \sin^2 \phi$, получаем

$$2\sin^2 \theta - 2\sin^4 \theta - (1 - \sin^2 \theta)f_z^2 = 0$$

или

$$2\sin^4 \theta - (2 + f_z^2)\sin^2 \theta + f_z^2 = 0. \quad (3-38)$$

Положив $u = \sin^2 \theta$, получим решения $\sin^2 \theta = f_z^2/2, 1$. Решение $\sin^2 \theta = 1$ отбросим, так как при подстановке его в (3-37) оно дает бесконечный результат. Следовательно,

$$\theta = \arcsin(\pm f_z / \sqrt{2}). \quad (3-39)$$

Подстановка в равенство (3-37) приводит к

$$\phi = \arcsin(\pm f_z / \sqrt{2 - f_z^2}). \quad (3-40)$$

Это показывает, что диапазон коэффициентов искажения равен $0 \leq f \leq 1$ ². Заметим далее, что каждый коэффициент искажения f_z порождает четыре возможных диметрических проекции.

На рис. 3-17 показаны диметрические проекции для разных коэффициентов искажения. Для каждого коэффициента была выбрана диметрическая проекция, соответствующая положительному вращению вокруг оси y и положительному вращению вокруг x .

На рис. 3-18 показаны четыре возможных диметрических проекции для коэффициента искажения $5/8$.

Рассмотрим конкретный пример.

¹ Могут быть использованы любые оси из трех.

² Отрицательные коэффициенты искажения не имеют смысла.

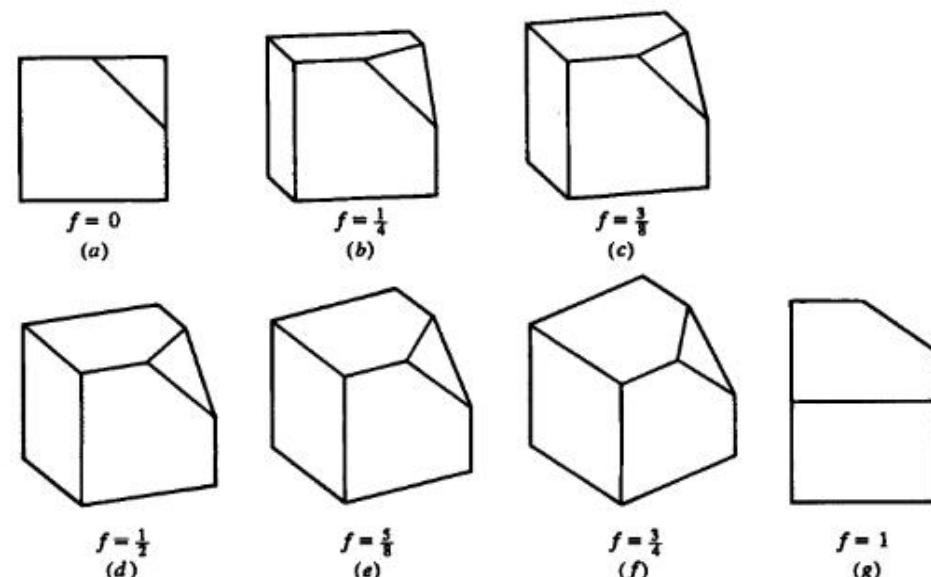


Рис. 3-17 Диметрические проекции для разных значений коэффициента искажения. (a) 0; (b) 1/4; (c) 3/8; (d) 1/2; (e) 5/8; (f) 3/4; (g) 1.

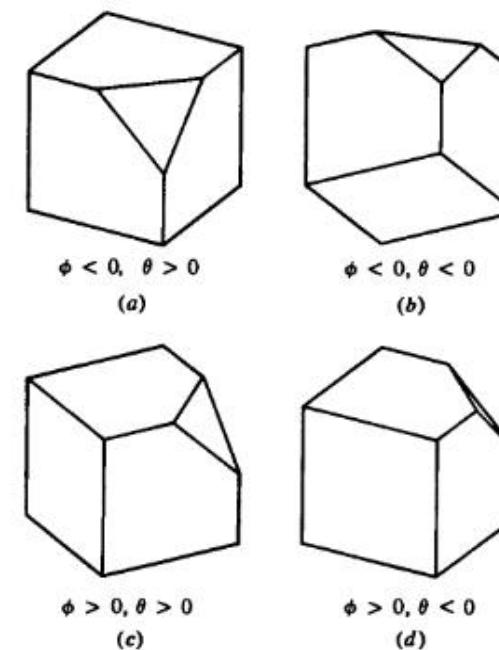


Рис. 3-18 Четыре возможных диметрических проекции для коэффициента искажения $5/8$ и углов поворотов $\phi = \pm 29.52^\circ, \theta = \pm 26.23^\circ$. (a) $\phi = -29.52^\circ, \theta = +26.23^\circ$; (b) $\phi = -29.52^\circ, \theta = -26.23^\circ$; (c) $\phi = +29.52^\circ, \theta = +26.23^\circ$; (d) $\phi = +29.52^\circ, \theta = -26.23^\circ$.

Пример 3-14 Диметрические проекции

Для куба с отсеченным углом построим диметрическую проекцию для коэффициента искажения по оси z , равного $1/2$.

Из равенства (3-39)

$$\begin{aligned}\theta &= \arcsin(\pm f_z / \sqrt{2}) \\ &= \arcsin(\pm 1/2\sqrt{2}) \\ &= \arcsin(\pm 0.35355) \\ &= \pm 20.705^\circ.\end{aligned}$$

Из равенства (3-40)

$$\begin{aligned}\phi &= \arcsin(\pm f_z / \sqrt{2 - f_z^2}) \\ &= \arcsin(\pm 1/2\sqrt{7/4}) \\ &= \arcsin(\pm 0.378) \\ &= \pm 22.208^\circ.\end{aligned}$$

Выбрав $\phi = +22.208^\circ$ и $\theta = +20.705^\circ$, из (3-32) получаем следующую матрицу диметрической проекции:

$$[T] = \begin{bmatrix} 0.926 & 0.134 & 0 & 0 \\ 0 & 0.935 & 0 & 0 \\ 0.378 & -0.327 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Вспоминая координатный вектор $[X]$ для куба с отсеченным углом (см. пример 3-13), получим координаты

$$[X^*] = [X][T] = \begin{bmatrix} 0.378 & -0.327 & 0 & 1 \\ 1.304 & -0.194 & 0 & 1 \\ 1.304 & 0.274 & 0 & 1 \\ 0.841 & 0.675 & 0 & 1 \\ 0.378 & 0.608 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.926 & 0.134 & 0 & 1 \\ 0.926 & 1.069 & 0 & 1 \\ 0 & 0.935 & 0 & 1 \\ 1.115 & 0.905 & 0 & 1 \end{bmatrix}.$$

Результат изображен на рис. 3-17 d.

Диметрическая проекция позволяет проводить измерения с одинаковым масштабным множителем по двум преобразованным главным осям. Измерение вдоль третьей оси требует другого масштабного множителя. Это может привести к путанице и ошибкам, если требуется точное масштабирование размеров спроектированного объекта. Изометрическая проекция решает эту проблему.

В изометрической проекции все три коэффициента искажения равны. Вспомним уравнения (3-34)–(3-36) и приравнивая (3-34) и (3-35), снова получим ра-

венство (3-37), т. е.

$$\sin^2 \phi = \frac{\sin^2 \theta}{1 - \sin^2 \theta}. \quad (3-37)$$

Приравнивание равенств (3-35) и (3-36) дает

$$\sin^2 \theta = \frac{1 - 2 \sin^2 \theta}{1 - \sin^2 \theta}. \quad (3-41)$$

Из (3-37) и (3-41) следует, что $\sin^2 \theta = 1/3$ или $\sin \theta = \pm \sqrt{1/3}$ и $\theta = \pm 35.26^\circ$. Тогда

$$\sin^2 \phi = \frac{1}{3} : \left(1 - \frac{1}{3}\right) = \frac{1}{2}$$

и $\phi = \pm 45^\circ$. Снова отметим, что существуют четыре возможных изометрических проекций. Это показано на рис. 3-19. Коэффициент искажения для изометрической проекции равен (см. (3-35))

$$f = \sqrt{\cos^2 \theta} = \sqrt{2/3} = 0.8165.$$

В самом деле, изометрическая проекция есть частный случай диметрической с $f_z = 0.8165$.

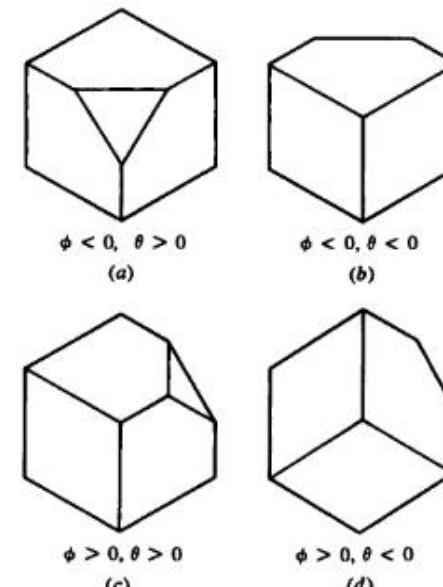


Рис. 3-19 Четыре возможных изометрических проекций с углами поворотов $\phi = \pm 45^\circ$, $\theta = \pm 35.26^\circ$. (a) $\phi = -45^\circ$, $\theta = +35.26^\circ$; (b) $\phi = -45^\circ$, $\theta = -35.26^\circ$; (c) $\phi = +45^\circ$, $\theta = +35.26^\circ$; (d) $\phi = +45^\circ$, $\theta = -35.26^\circ$.

При построении изометрических проекций вручную важен угол, который составляет проекция оси x с горизонталью. Преобразование единичного вектора вдоль оси x с помощью матрицы изометрической проекции дает

$$[U_x^*] = [1 \ 0 \ 0 \ 1] \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & 0 \\ 0 & \cos \theta & 0 & 0 \\ \sin \phi & -\cos \phi \sin \theta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ = [\cos \phi \ \sin \phi \sin \theta \ 0 \ 1].$$

Тогда угол между проекцией оси x и горизонталью равен

$$\operatorname{tg} \alpha = \frac{y_x^*}{x_x^*} = \frac{\sin \phi \sin \theta}{\cos \phi} = \pm \sin \theta, \quad (3-42)$$

поскольку $\sin \phi = \cos \phi$ для $\phi = 45^\circ$. Следовательно,

$$\alpha = \operatorname{arctg}(\pm \sin 35.26439^\circ) = \pm 30^\circ.$$

Для построения изометрических проекций вручную обычно используется пластмассовый прямоугольный треугольник с углами 30° и 60° . Ниже приводится пример.

Пример 3-15 Изометрическая проекция

Снова рассмотрим куб с отсеченным углом (см. пример 3-13) и построим изометрическую проекцию для $\phi = -45^\circ$ и $\theta = +35.26439^\circ$. Из равенства (3-32) получаем, что преобразование изометрического проецирования имеет вид:

$$[T] = \begin{bmatrix} 0.707 & -0.408 & 0 & 0 \\ 0 & 0.816 & 0 & 0 \\ -0.707 & -0.408 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Вспоминая координаты $[X]$, получим

$$[X^*] = [X][T] = \begin{bmatrix} -0.707 & -0.408 & 0 & 1 \\ 0 & -0.816 & 0 & 1 \\ 0 & -0.408 & 0 & 1 \\ -0.354 & 0.204 & 0 & 1 \\ -0.707 & 0.408 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.707 & -0.408 & 0 & 1 \\ 0.707 & 0.408 & 0 & 1 \\ 0 & 0.816 & 0 & 1 \\ 0.354 & 0.204 & 0 & 1 \end{bmatrix}.$$

Результат изображен на рис. 3-19 а.

3-14 КОСОУГОЛЬНЫЕ ПРОЕКЦИИ

В противоположность ортографическим и аксонометрическим проекциям, для которых проекторы перпендикулярны плоскости проекции, косоугольная проекция формируется параллельными проекторами с центром, лежащим в бесконечности, расположенным под косым углом к плоскости проекции. Общая схема проекции изображена на рис. 3-20.

Косоугольные проекции показывают общую трехмерную форму объекта. Однако истинные размер и форма изображаются только для граней объекта, расположенных параллельно плоскости проекции, т. е. углы и длины сохраняются только для таких граней. В самом деле, косоугольная проекция этих граней эквивалентна ортографическому виду спереди. Границы, не параллельные плоскости проекции, подвергаются искажению.

Особый интерес представляют две косоугольные проекции — кавалье и кабине. Проекция кавалье получается, когда угол между проекторами и плоскостью проекции составляет 45° . В этой проекции коэффициенты искажения для всех трех главных направлений одинаковы. Результат этой проекции выглядит неестественно утолщенным. Для «коррекции» этого недостатка используется проекция кабине.

Проекцией кабине называется такая косоугольная проекция, у которой коэффициент искажения для ребер, перпендикулярных плоскости проекции, равен $1/2$. Как будет показано ниже, для проекции кабине угол между проекторами и плоскостью проекции составляет $\operatorname{arcctg}(1/2) = 63.43^\circ$.

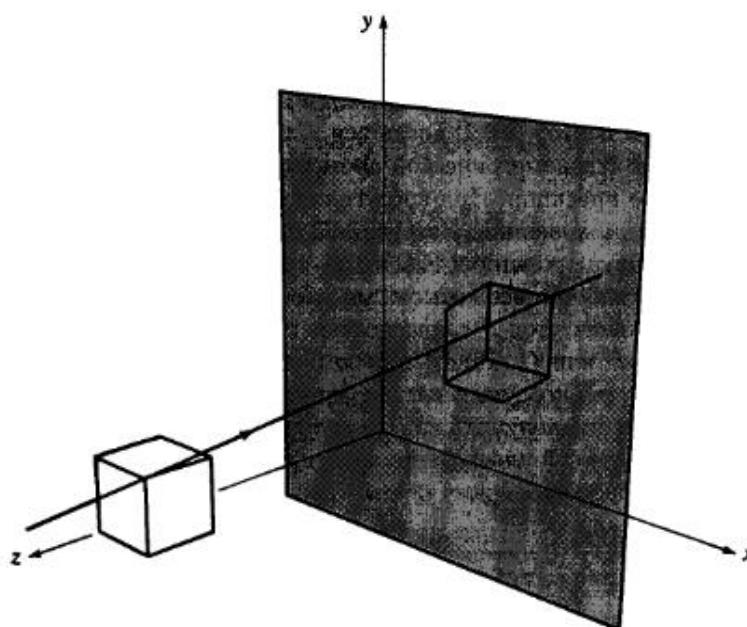


Рис. 3-20 Косоугольная проекция.

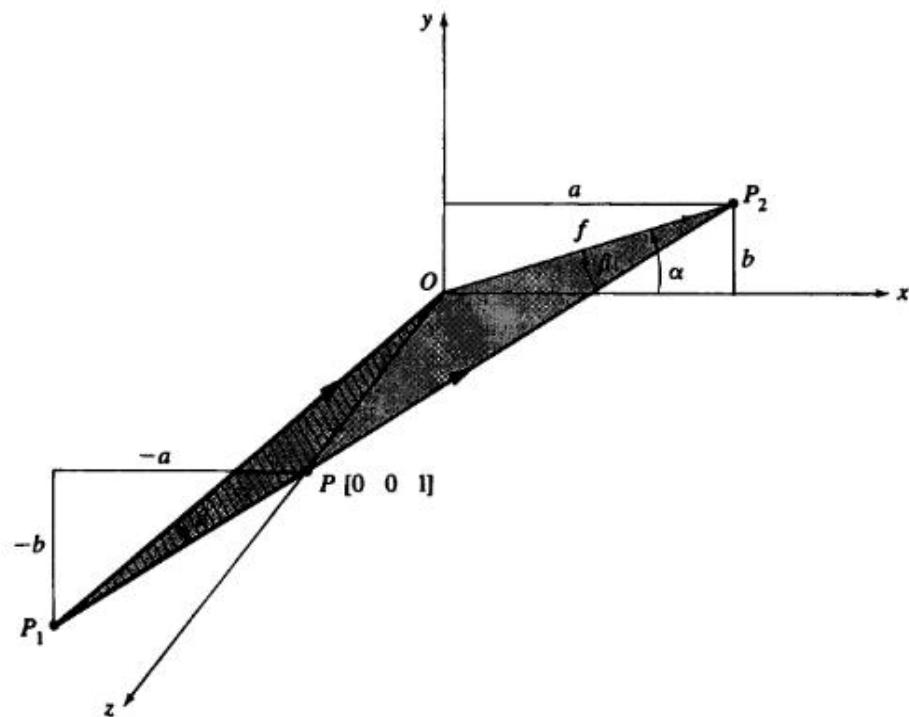


Рис. 3-21 Построение косоугольной проекции.

Чтобы построить матрицу преобразования для косоугольной проекции, рассмотрим единичный вектор $[0 \ 0 \ 1]$ вдоль оси z , показанный на рис. 3-21. Для ортографической или аксонометрической проекции на плоскость $z = 0$ вектор PO задает направление проекции. При косоугольной проекции проекторы составляют угол с плоскостью проекции. На рис. 3-21 показаны типичные косоугольные проекторы P_1O и PP_2 . Проекторы P_1O и PP_2 образуют угол β с плоскостью проекции $z = 0$. Заметим, что все возможные проекторы, проходящие через точку P или O и образующие угол β с плоскостью $z = 0$, лежат на поверхности конуса с вершиной в P или O . Таким образом, для заданного угла β существует бесконечное количество косоугольных проекций.

Проектор P_1O можно получить из PO с помощью переноса на $[-a - b 1]$ точки P в точку P_1 . В двумерной плоскости, проходящей через P перпендикулярно оси z , (3×3) -матрица преобразования равна

$$[T'] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -a & -b & 1 \end{bmatrix}.$$

В трехмерном пространстве это двумерное преобразование эквивалентно сдвигу вектора PO в направлениях x и y . Для этого необходимо преобразование

$$[T''] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -a & -b & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Проектирование на плоскость $z = 0$ дает

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -a & -b & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Из рис. 3-21 получаем, что

$$\begin{aligned} a &= f \cos \alpha, \\ b &= f \sin \alpha, \end{aligned}$$

где f — длина спроцированного единичного вектора на оси z , т. е. коэффициент искажения, а α — угол между горизонталью и спроцированной осью z . Из рис. 3-21 также ясно, что β — угол между косыми проекторами и плоскостью проекции равен

$$\beta = \operatorname{arcctg}(f). \quad (3-43)$$

Таким образом, преобразование для косоугольной проекции имеет вид:

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -f \cos \alpha & -f \sin \alpha & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-44)$$

При $f = 0$, $\beta = 90^\circ$ получаем ортографическую проекцию. Если $f = 1$, то не подвергаются искажению ребра, перпендикулярные плоскости проекции. А это является условием проекции кавалье. Из равенства (3-43) имеем:

$$\beta = \operatorname{arcctg}(1) = 45^\circ.$$

Заметим, что в проекции кавалье α является все еще свободным параметром. На рис. 3-22 показаны проекции кавалье для некоторых значений α . Наиболее часто используются значения α , равные 30° и 45° . Так же применяется значение $180^\circ - \alpha$.

Проекцию кавалье можно получить при коэффициенте искажения $f = 1/2$. Отсюда

$$\beta = \operatorname{arcctg}(1/2) = 63.435^\circ.$$

В этом случае снова угол α является переменной величиной, как это показано на рис. 3-23. Наиболее часто встречаются значения 30° и 45° , применяется также

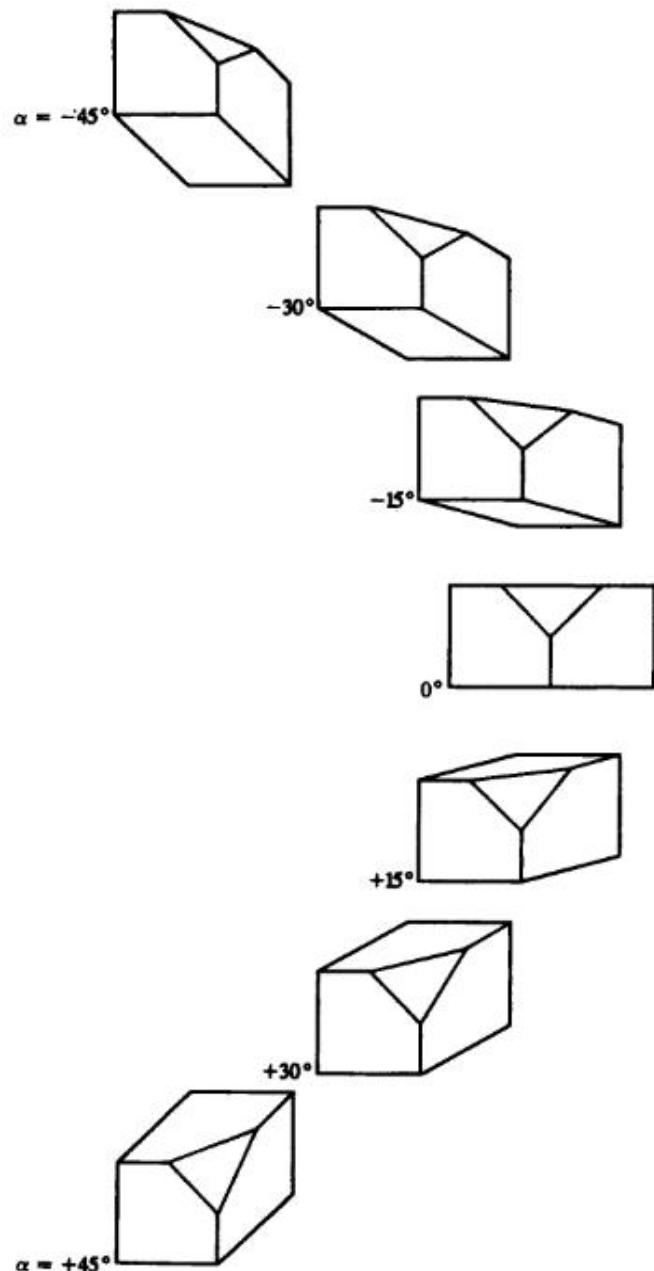


Рис. 3-22 Проекции кавалье. Сверху вниз угол α изменяется от -45° до $+45^\circ$ с интервалом 15° , угол $\beta = 45^\circ$.

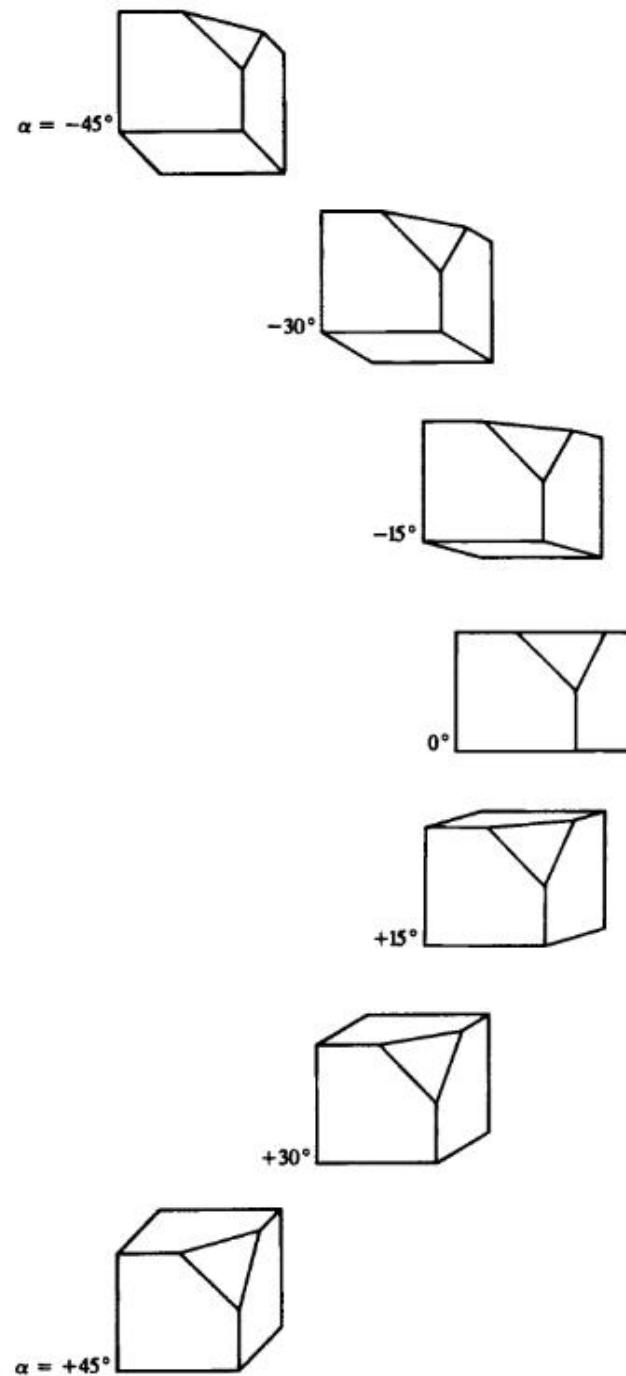


Рис. 3-23 Проекции кабине. Сверху вниз угол α изменяется от -45° до $+45^\circ$ с интервалом 15° , коэффициент искажения $f = 0.5$.

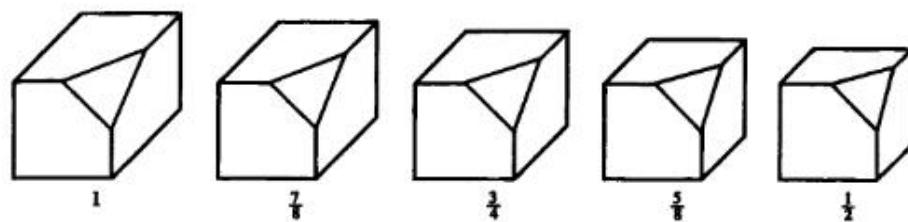


Рис. 3-24 Косоугольные проекции. Слева направо $f = 1, 7/8, 3/4, 5/8, 1/2$ при $\alpha = 45^\circ$.

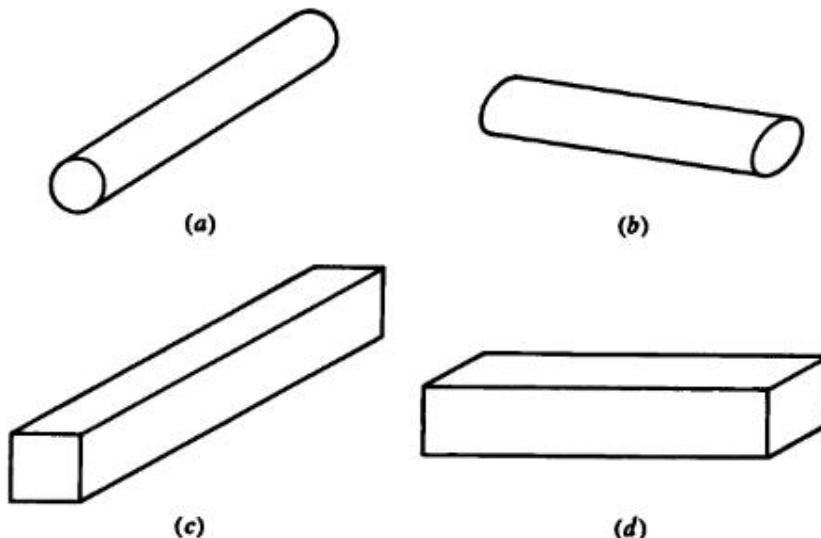


Рис. 3-25 Искажение, возникающее в косоугольных проекциях, $f = 5/8$, $\alpha = 45^\circ$.
 (a) Круглая грань параллельна плоскости проекции; (b) круглая грань перпендикулярна плоскости проекции; (c) длинная сторона перпендикулярна плоскости проекции;
 (d) длинная сторона параллельна плоскости проекции.

значение $180^\circ - \alpha$. На рис. 3-24 изображены косоугольные проекции для коэффициентов искажения $f = 1, 7/8, 3/4, 5/8, 1/2$ с углом $\alpha = 45^\circ$.

Поскольку изображается истинная форма одной грани, косоугольные проекции особенно подходят для иллюстрации объектов с круглыми или иными искривленными гранями. Такие грани должны быть параллельны плоскости проекции, чтобы избежать нежелательных искажений. Так же, как и в случае параллельных проекций, объекты с одним измерением, существенно превосходящим другие, подвергаются значительному искажению, если только это измерение не параллельно плоскости проекции. Такие эффекты показаны на рис. 3-25.

Ниже приводится подробный пример.

Пример 3-16 Косоугольные проекции

Построим проекции кавалье и кабине для куба с одним отсеченным углом (см. пример 3-13).

Вспомнив, что проекция кавалье – это косоугольная проекция с $\beta = 45^\circ$, т. е. коэффициент искажения $f = 1$, и выбрав угол горизонтального наклона $\alpha = 30^\circ$, из (3-44) получим матрицу преобразования

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -f \cos \alpha & -f \sin \alpha & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -0.866 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразуем координаты $[X]$ куба с отсеченным углом (см. пример 3-13)

$$[X^*] = [X][T] = \begin{bmatrix} -0.866 & -0.5 & 0 & 1 \\ 0.134 & -0.5 & 0 & 1 \\ 0.134 & 0 & 0 & 1 \\ -0.366 & 0.5 & 0 & 1 \\ -0.866 & 0.5 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0.567 & 0.75 & 0 & 1 \end{bmatrix}.$$

Результат изображен на рис. 3-22.

Обратившись теперь к проекции кабине и вспомнив, что коэффициент искажения равен $1/2$, из равенства (3-43) получим

$$\beta = \operatorname{arcctg}(1/2) = \operatorname{arctg}(2) = 63.435^\circ.$$

Снова выбрав $\alpha = 30^\circ$, из (3-44), имеем:

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -f \cos \alpha & -f \sin \alpha & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -0.433 & -0.25 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразованные координаты проекции кабине для этого куба равны

$$[X^*] = [X][T] = \begin{bmatrix} -0.433 & -0.25 & 0 & 1 \\ 0.567 & -0.25 & 0 & 1 \\ 0.567 & 0.25 & 0 & 1 \\ 0.067 & 0.75 & 0 & 1 \\ -0.433 & 0.75 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0.783 & 0.875 & 0 & 1 \end{bmatrix}.$$

Этот результат изображен на рис. 3-23.

Отметим, что и для проекции кавалье, и для проекции кабине треугольная грань не изображается с правильными размерами или формой, так как она не параллельна плоскости проекции ($z = 0$).

3-15 ПЕРСПЕКТИВНЫЕ ПРЕОБРАЗОВАНИЯ

Перспективное преобразование имеет место, когда не равен нулю любой из первых трех элементов четвертого столбца обобщенной (4×4) -матрицы преобразования однородных координат. Как было упомянуто ранее (см. разд. 3-11), перспективное преобразование — это преобразование одного трехмерного пространства в другое. В отличие от обсуждавшихся ранее параллельных преобразований, в данном случае параллельные прямые сходятся, размер объекта уменьшается с увеличением расстояния до центра проекции, и происходит неоднородное искашение линий объекта, зависящее от ориентации и расстояния от объекта до центра проекции. Все это помогает нашему восприятию глубины, но не сохраняет форму объекта.

Одноточечное перспективное преобразование задается равенством

$$[x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ rz + 1]. \quad (3-45)$$

Здесь $h = rz + 1 \neq 1$. Обычные координаты получаются делением на h :

$$[x^* \ y^* \ z^* \ 1] = \left[\frac{x}{rz+1} \ \frac{y}{rz+1} \ \frac{z}{rz+1} \ 1 \right]. \quad (3-46)$$

Перспективное проецирование на некоторую двумерную видовую плоскость можно получить, объединив ортографическую проекцию с перспективным преобразованием. Например, перспективное проецирование на плоскость $z = 0$ выполняется с помощью преобразований

$$\begin{aligned} [T] &= [P_r] [P_z] = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (3-47)$$

и

$$[x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ 0 \ rz + 1]. \quad (3-48)$$

Обычные координаты равны

$$[x^* \ y^* \ z^* \ 1] = \left[\frac{x}{rz+1} \ \frac{y}{rz+1} \ 0 \ 1 \right]. \quad (3-49)$$

Чтобы показать, что равенство (3-47) выполняет перспективное проецирование на плоскость $z = 0$, рассмотрим рис. 3-26 с геометрическими построениями для перспективной проекции трехмерной точки P на плоскость $z = z^* = 0$ в точку P^*

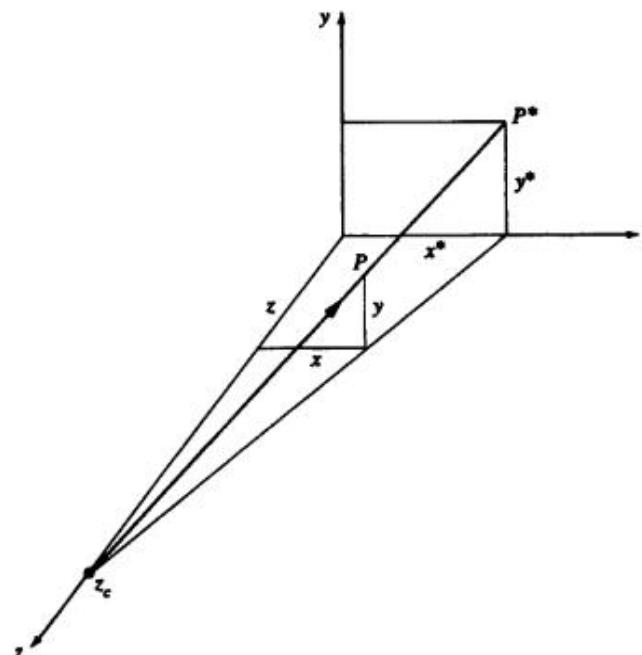


Рис. 3-26 Перспективная проекция точки.

и с центром проекции, лежащим в z_c на оси z . Координаты точки проекции P^* можно получить, используя подобие треугольников. Из рис. 3-26 следует:

$$\frac{x^*}{z_c} = \frac{x}{z_c - z}$$

или

$$x^* = x / \left(1 - \frac{z}{z_c} \right)$$

и

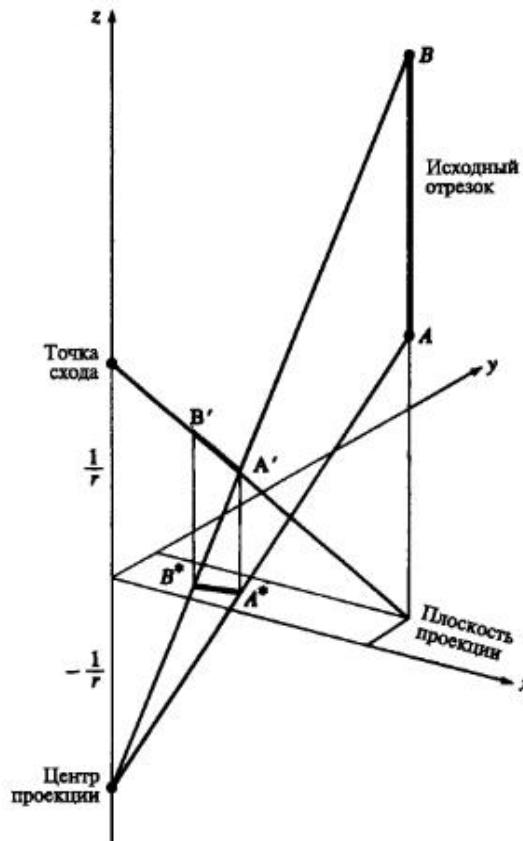
$$\frac{y^*}{\sqrt{x^{*2} + z_c^2}} = \frac{y}{\sqrt{x^2 + (z_c - z)^2}}$$

или

$$y^* = \frac{y}{1 - \frac{z}{z_c}},$$

z^* равно, конечно, нулю.

Полагая $r = -1/z_c$, получим результаты, аналогичные полученным с помощью (3-47). Таким образом, уравнение (3-47) осуществляет перспективное проецирование на плоскость $z = 0$ с центром проекции в точке $(-1/r)$ на оси z . Заметим,

Рис. 3-27 Проекция прямой, параллельной оси z .

что при приближении z_c к бесконечности, r приближается к нулю и в результате получаем аксонометрическую проекцию на плоскость $z = 0$. Заметим далее, что на точки, лежащие в плоскости проекции, т. е. $z = 0$, перспективное преобразование не действует. Отметим также, что начало координат ($x = y = z = 0$) остается неизменным. Следовательно, если плоскость проекции ($z = 0$) проходит через объект, то эта часть объекта изображается с правильным размером и формой. Все другие части объекта искажаются.

Чтобы лучше понять действие перспективного преобразования, рассмотрим рис. 3-27. На нем показано перспективное проецирование на плоскость $z = 0$ отрезка AB , параллельного оси z , в отрезок $A'B'$ на плоскости $z = 0$ с центром проекции, расположенным в точке $-1/r$ на оси z . Преобразование можно разделить на два этапа (см. (3-47)). На первом этапе отрезок AB отображается в отрезок $A'B'$ (см. рис. 3-27). Затем с помощью ортографического проецирования отрезок $A'B'$ в трехмерном пространстве отображается в отрезок A^*B^* на плоскости $z = 0$. Центр проекции расположен в бесконечности.

Исследование рис. 3-27 показывает, что прямые $A'B'$ и AB пересекают плоскость $z = 0$ в одной и той же точке. Прямая $A'B'$ также пересекает ось z в точке $z = +1/r$. Далее, перспективное преобразование (см. (3-45) и (3-46)) отображает расположенную в бесконечности точку пересечения параллельных прямых AB и оси z в конечную точку $z = 1/r$ на оси z . Эта точка называется точкой схода¹. Заметим, что точка схода лежит на том же расстоянии от плоскости проекции, что и центр проекции, только с противоположной стороны от плоскости, например, если $z = 0$ есть плоскость проекции, а центр проекции находится в $z = -1/r$, тогда точка схода находится в $z = +1/r$.

Чтобы подтвердить это наблюдение, рассмотрим перспективное преобразование точки, находящейся в бесконечности на оси $+z$, т. е.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} = [0 \ 0 \ 1 \ r]. \quad (3-50)$$

Соответствующая ей точка $[x^* \ y^* \ z^* \ 1] = [0 \ 0 \ 1/r \ 1]$ теперь является конечной точкой на положительной оси z . Это означает, что все полубесконечное положительное пространство ($0 \leq z \leq \infty$) отображается в ограниченную область $0 \leq z^* \leq 1/r$. Далее, все прямые, параллельные оси z , теперь проходят через точку $[0 \ 0 \ 1/r \ 1]$ – точку схода.

Прежде чем перейти к примерам, для полноты изложения укажем одноточечные перспективные преобразования с центром проекции и точкой схода, расположенными на осях x и y . Одноточечное перспективное преобразование

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & p \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ (px + 1)] \quad (3-51)$$

с обычными координатами

$$[x^* \ y^* \ z^* \ 1] = \left[\frac{x}{px + 1} \ \frac{y}{px + 1} \ \frac{z}{px + 1} \ 1 \right] \quad (3-52)$$

имеет центр проекции $[-1/p \ 0 \ 0 \ 1]$ и точку схода, расположенную на оси x в $[1/p \ 0 \ 0 \ 1]$.

Одноточечное перспективное преобразование

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & q \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ (qy + 1)] \quad (3-53)$$

¹ Интуитивно точка схода – это такая точка «вдали», в которой параллельные прямые «каружутся» сходящимися и исчезающими. В качестве примера приведем прямолинейную, уходящую вдаль железнодорожную колею.

с обычными координатами

$$\begin{bmatrix} x^* & y^* & z^* & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 \\ qy+1 & qy+1 & qy+1 & 1 \end{bmatrix} \quad (3-54)$$

имеет центр проекции $[0 \ -1/q \ 0 \ 1]$ и точку схода, расположенную на оси y в $[0 \ 1/q \ 0 \ 1]$.

Пример 3-17 Перспективное преобразование прямой, параллельной оси z

Рассмотрим отрезок AB на рис. 3-27, параллельный оси z , с концевыми точками $A[3 \ 2 \ 4 \ 1]$ и $B[3 \ 2 \ 8 \ 1]$. Выполним перспективное проецирование на плоскость $z = 0$ с центром проекции в точке $z = -2$. Перспективное преобразование AB в $A'B'$ при $r = 0.5$ равно

$$\begin{aligned} A[3 \ 2 \ 4 \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} &= \begin{bmatrix} 3 & 2 & 4 & 3 \\ 3 & 2 & 8 & 5 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0.667 & 1.333 & 1 \\ 0.6 & 0.4 & 1.6 & 1 \end{bmatrix} A' \\ &B[3 \ 2 \ 8 \ 1] \end{aligned}$$

$$B'.$$

Параметрическое уравнение отрезка $A'B'$

$$P(t) = [A'] + [B' - A']t, \quad 0 \leq t \leq 1,$$

или $P(t) = [1 \ 0.667 \ 1.333 \ 1] + [-0.4 \ -0.267 \ 0.267 \ 0]t$.

Пересечение этого отрезка с плоскостями $x = 0$, $y = 0$ и $z = 0$ дает

$$x(t) = 0 = 1 - 0.4t \rightarrow t = 2.50$$

$$y(t) = 0 = 0.667 - 0.267t \rightarrow t = 2.50$$

$$z(t) = 0 = 1.333 + 0.267t \rightarrow t = -5.0.$$

Подставив значение $t = 2.5$ в параметрическое уравнение отрезка $A'B'$, получим

$$z(2.5) = 1.333 + (0.267)(2.5) = 2.0,$$

что является пересечением отрезка $A'B'$ с осью z в точке $z = +1/r$, точке схода. Теперь подстановка в уравнения для x - и y -компонент значения $t = -5.0$ дает пересечение с плоскостью $z = 0$, т. е.

$$x(-5.0) = 1 - (0.4)(-5.0) = 3.0$$

$$y(-5.0) = 0.667 - (0.267)(-5.0) = 2.0,$$

что совпадает с пересечением прямой AB с плоскостью $z = 0$. Проекция отрезка $A'B'$ в отрезок A^*B^* на плоскости $z = 0$ вычисляется следующим образом

$$\begin{aligned} A' \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0.667 & 0 & 1 \\ 0.6 & 0.4 & 0 & 1 \end{bmatrix} A^* \\ B' \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0.667 & 0 & 1 \\ 0.6 & 0.4 & 0 & 1 \end{bmatrix} B^* \end{aligned}$$

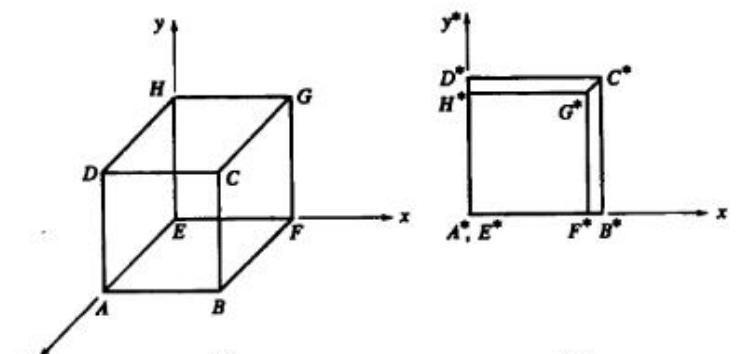
Ниже приводится пример с простым кубом.

Пример 3-18 Одноточечное перспективное преобразование куба

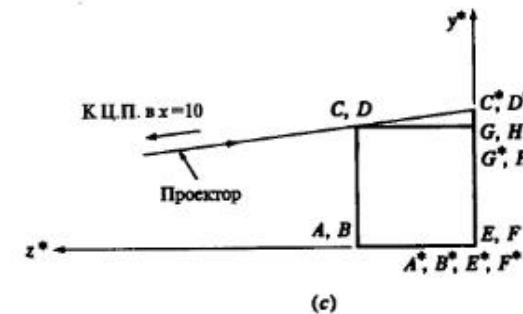
Выполним перспективное проецирование на плоскость $z = 0$ единичного куба, изображенного на рис. 3-28 а, с центром проекции в точке $z_c = 10$ на оси z .

Одноточечный перспективный множитель r равен

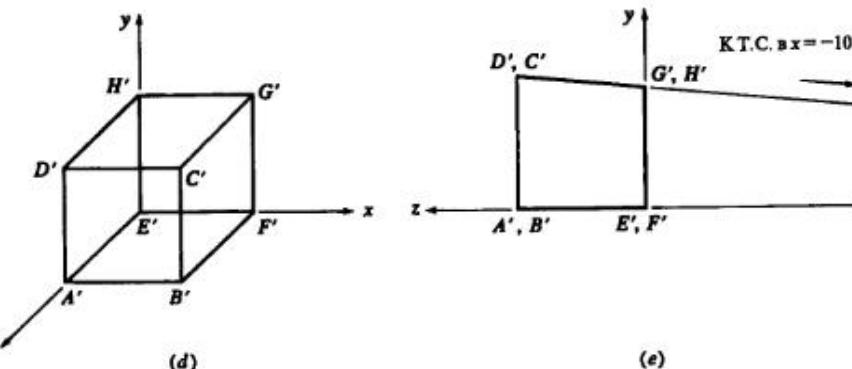
$$r = -1/z_c = -1/10 = -0.1.$$



(a) (b)



(c)



(d) (e)

Рис. 3-28 Одноточечная перспективная проекция единичного куба.

Из уравнения (3-48) получаем, что

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\begin{aligned} [X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 0 & 0 & 0.9 \\ 1 & 0 & 0 & 0.9 \\ 1 & 1 & 0 & 0.9 \\ 0 & 1 & 0 & 0.9 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1.11 & 0 & 0 & 1 \\ 1.11 & 1.11 & 0 & 1 \\ 0 & 1.11 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результат изображен на рис. 3-28 б. Отметим, что, поскольку центр проекции находится на положительной оси z , проекция передней грани $ABCD$ куба больше проекции задней грани. Почему так происходит, показано на рис. 3-28 с, на котором изображена параллельная проекция исходного куба на плоскость $x = 0$.

Отметим также, что, поскольку точка схода лежит на оси z , прямая C^*G^* на рис. 3-28 б проходит через начало координат.

В качестве альтернативного метода, эквивалентного первому, можно было выполнить перспективное преобразование и получить искаженный объект в трехмерном пространстве, а затем ортографически спроектировать результат на некоторую плоскость. Искаженный объект получают следующим образом:

$$\begin{aligned} [X'] &= [X][P_r] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 0 & 1 & 0.9 \\ 1 & 0 & 1 & 0.9 \\ 1 & 1 & 1 & 0.9 \\ 0 & 1 & 1 & 0.9 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1.11 & 1 \\ 1.11 & 0 & 1.11 & 1 \\ 1.11 & 1.11 & 1.11 & 1 \\ 0 & 1.11 & 1.11 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результат, полученный с помощью косоугольной проекции показан на рис. 3-28 д. Заметим, что «передняя» грань ($A'B'C'D'$) больше, чем «задняя» грань ($E'F'G'H'$). Последующее ортографическое проецирование на плоскость $z = 0$ дает тот же самый результат $[X^*]$, что был получен ранее и который изображен на рис. 3-28 с.

На рис. 3-28 е, на котором представлена ортографическая проекция на плоскость $x = 0$ искаженного объекта с рис. 3-28 д, показано, что ребра куба, ранее параллельные оси z , теперь сходятся к точке схода $z = -10$.

Рис. 3-28 б не передает трехмерности куба. Как показано в следующем примере, более удовлетворительный результат можно получить центрированием куба.

Пример 3-19 Одноточечное перспективное преобразование центрированного куба

Изображенный на рис. 3-28 а куб может быть центрирован на оси z путем его переноса на $-1/2$ вдоль направлений x и y . Результирующее преобразование

$$\begin{aligned} [T] &= [Tr_{xy}][P_{rz}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -0.5 & -0.5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.1 \\ -0.5 & -0.5 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Сдвинутый куб изображен на рис. 3-29 а.

Преобразованные обычные координаты равны

$$\begin{aligned} [X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.1 \\ -0.5 & -0.5 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} -0.5 & -0.5 & 0 & 0.9 \\ 0.5 & -0.5 & 0 & 0.9 \\ 0.5 & 0.5 & 0 & 0.9 \\ -0.5 & 0.5 & 0 & 0.9 \\ -0.5 & -0.5 & 0 & 1 \\ 0.5 & -0.5 & 0 & 1 \\ 0.5 & 0.5 & 0 & 1 \\ -0.5 & 0.5 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -0.56 & -0.56 & 0 & 1 \\ 0.56 & -0.56 & 0 & 1 \\ 0.56 & 0.56 & 0 & 1 \\ -0.56 & 0.56 & 0 & 1 \\ -0.5 & -0.5 & 0 & 1 \\ 0.5 & -0.5 & 0 & 1 \\ 0.5 & 0.5 & 0 & 1 \\ -0.5 & 0.5 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результат изображен на рис. 3-29 б. Заметим, что ранее параллельные оси z прямые, соединяющие углы передней и задней граней, теперь сходятся к оси z ($x = 0$, $y = 0$).

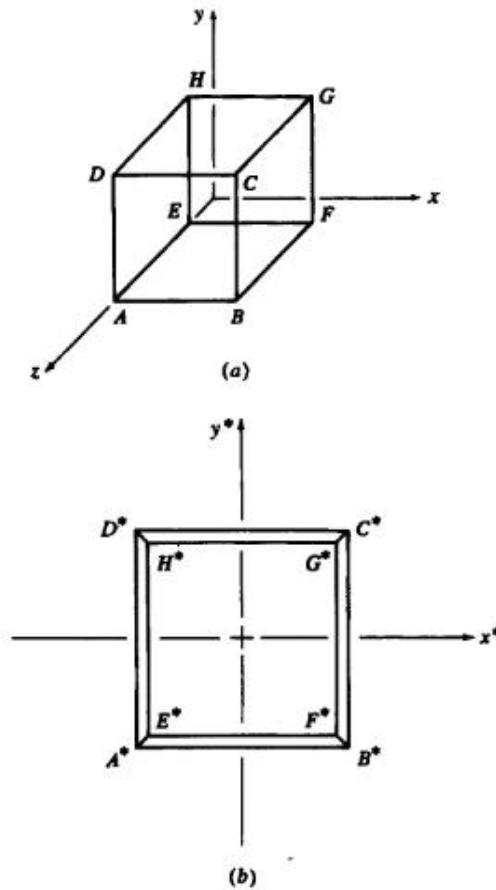


Рис. 3-29 Одноточечная перспективная проекция центрированного единичного куба.

К сожалению, результирующее изображение все еще не дает адекватного восприятия трехмерной формы объекта. Поэтому мы обратимся к более сложным перспективным преобразованиям.

Если в четвертом столбце (4×4)-матрицы преобразования два элемента из первых трех не равны нулю, то такое преобразование называется двуточечным перспективным преобразованием. Двуточечное перспективное преобразование

$$[x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & p \\ 0 & 1 & 0 & q \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ (px + qy + 1)] \quad (3-55)$$

с обычными координатами

$$[x^* \ y^* \ z^* \ 1] = \left[\frac{x}{px + qy + 1} \quad \frac{y}{px + qy + 1} \quad \frac{z}{px + qy + 1} \quad 1 \right] \quad (3-56)$$

имеет два центра проекции: первый на оси x в точке $[-1/p \ 0 \ 0 \ 1]$ и второй на оси y в точке $[0 \ -1/q \ 0 \ 1]$, и две точки схода: на оси x в точке $[1/p \ 0 \ 0 \ 1]$ и на оси y в точке $[0 \ 1/q \ 0 \ 1]$. Заметим, что заданное уравнением (3-55) двуточечное преобразование можно получить объединением двух одноточечных. Конкретнее,

$$[P_{pq}] = [P_p][P_q] = [P_q][P_p],$$

где $[P_{pq}]$ задается уравнением (3-55), $[P_p]$ — уравнением (3-53) и $[P_q]$ — уравнением (3-51). Более подробно двуточечная перспективная проекция рассмотрена в следующем примере.

Пример 3-20 Двуточечные перспективные преобразования

Снова рассмотрим куб, описанный в примере 3-18. Построим двуточечную перспективную проекцию этого куба на плоскость $z = 0$ для центров проекции, находящихся в точках $x = -10$ и $y = -10$. Нужное преобразование получим путем объединения (3-55) и (3-27). А именно,

$$\begin{aligned} [T] = [P_{pq}][P_z] &= \begin{bmatrix} 1 & 0 & 0 & p \\ 0 & 1 & 0 & q \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & p \\ 0 & 1 & 0 & q \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Здесь p и q равны

$$p = -1/(-10) = 0.1, \quad q = -1/(-10) = 0.1.$$

Преобразованные координаты куба имеют вид:

$$\begin{aligned} [X^*] = [X][T] &= \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1.1 \\ 1 & 1 & 0 & 1.2 \\ 0 & 1 & 0 & 1.1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1.1 \\ 1 & 1 & 0 & 1.2 \\ 0 & 1 & 0 & 1.1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0.909 & 0 & 0 & 1 \\ 0.833 & 0.833 & 0 & 1 \\ 0 & 0.909 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.909 & 0 & 0 & 1 \\ 0.833 & 0.833 & 0 & 1 \\ 0 & 0.909 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результаты изображены на рис. 3-30 а. Две точки схода находятся в $x = 10$ и $y = 10$.

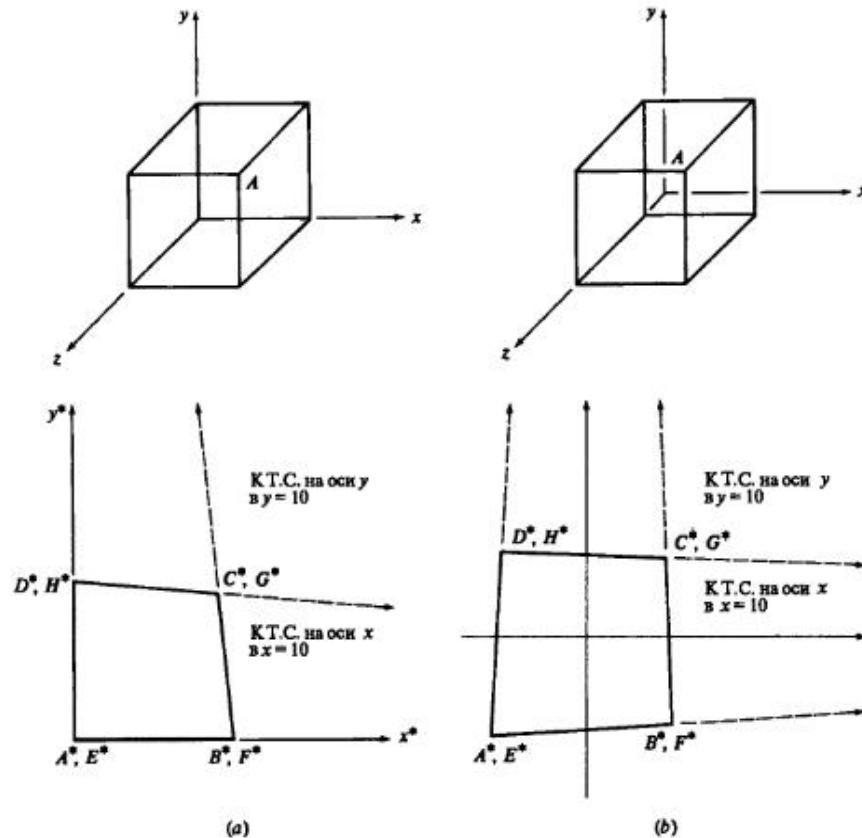


Рис. 3-30 Двухточечные перспективные проекции. (a) Нецентрированного куба; (b) центрированного куба.

Центрирование куба на оси z с помощью переноса вдоль x и y на -0.5 , аналогичное проделанному в примере 3-18, приведет к следующей общей матрице преобразования

$$\begin{aligned} [T] &= [T_{xy}] [P_{pq}] [P_z] = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -0.5 & -0.5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 0 & 0 \\ -0.5 & -0.5 & 0 & 0.9 \end{bmatrix}, \end{aligned}$$

где предполагается проецирование на плоскость $z = 0$. Заметим, что в этом случае общий масштабирующий множитель больше не равен единице (см.

(3-4)), т. е. происходит очевидное масштабирование куба, вызванное его перемещением. Преобразованные координаты равны

$$\begin{aligned} [X^*] &= [X] [T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 0 & 0 \\ -0.5 & -0.5 & 0 & 0.9 \end{bmatrix} = \\ &= \begin{bmatrix} -0.5 & -0.5 & 0 & 0.9 \\ 0.5 & -0.5 & 0 & 1 \\ 0.5 & 0.5 & 0 & 1.1 \\ -0.5 & 0.5 & 0 & 1 \\ -0.5 & -0.5 & 0 & 0.9 \\ 0.5 & -0.5 & 0 & 1 \\ 0.5 & 0.5 & 0 & 1.1 \\ -0.5 & 0.5 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -0.56 & -0.56 & 0 & 1 \\ 0.5 & -0.5 & 0 & 1 \\ 0.46 & 0.46 & 0 & 1 \\ -0.5 & 0.5 & 0 & 1 \\ -0.56 & -0.56 & 0 & 1 \\ 0.5 & -0.5 & 0 & 1 \\ 0.46 & 0.46 & 0 & 1 \\ -0.5 & 0.5 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результаты изображены на рис. 3-30 б.

И снова полученное изображение не дает адекватного восприятия трехмерной формы объекта. Поэтому обратимся к трехточечным перспективным преобразованиям.

Трехточечная перспектива получается, если не равны нулю три первых элемента четвертого столбца (4×4)-матрицы преобразования. Это трехточечное перспективное преобразование

$$[x \ y \ z \ 1] = \begin{bmatrix} 1 & 0 & 0 & p \\ 0 & 1 & 0 & q \\ 0 & 0 & 1 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ z \ (px + qy + rz + 1)] \quad (3-57)$$

с обычными координатами

$$[x^* \ y^* \ z^* \ 1] = \left[\frac{x}{px + qy + rz + 1} \ \frac{y}{px + qy + rz + 1} \ \frac{z}{px + qy + rz + 1} \ 1 \right] \quad (3-58)$$

имеет три центра проекции: на оси x в точке $[-1/p \ 0 \ 0 \ 1]$, на оси y в $[0 \ -1/q \ 0 \ 1]$ и на оси z в $[0 \ 0 \ -1/r \ 1]$, а также три точки схода: на оси x в $[1/p \ 0 \ 0 \ 1]$, на оси y в $[0 \ 1/q \ 0 \ 1]$ и на оси z в $[0 \ 0 \ 1/r \ 1]$.

Снова заметим, что трехточечное перспективное преобразование, заданное равенством (3-57), может быть получено конкатенацией трех одноточечных перспективных преобразований, по одному на каждую координатную ось. Создание трехточечной перспективы иллюстрируется на следующем примере.

Пример 3-21 Трехточечное перспективное преобразование

Для описанного в примере 3-18 куба рассмотрим проекцию на плоскость $z = 0$ после применения трехточечного перспективного преобразования. Центры проекции находятся в точках $x = -10, y = -10$ и $z = 10$. Точки схода находятся в $x = 10, y = 10$ и $z = -10$. Матрица преобразования равна

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразованные координаты куба

$$\begin{aligned} [X^*] = [X][T] &= \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0.1 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 0 & -0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0 & 0 & 0 & 0.9 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1.1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1.1 \\ 1 & 1 & 0 & 1.2 \\ 0 & 1 & 0 & 1.1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0.909 & 0.909 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.909 & 0 & 0 & 1 \\ 0.833 & 0.833 & 0 & 1 \\ 0 & 0.909 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результат показан на рис. 3-31 б. Искаженный в результате перспективного преобразования объект изображен на рис. 3-31 с. Отметим, что ребра на рисунке сходятся.

И снова получившийся вид недостаточно информативен, хотя и математически корректен. Соответствующие методы генерации перспективных видов обсуждаются в разд. 3-16.

3-16 МЕТОДЫ СОЗДАНИЯ ПЕРСПЕКТИВНЫХ ВИДОВ

Предложенные в предыдущем разделе виды перспективной проекции были неинформативны, так как во всех случаях из каждого центра проекции была видна только одна грань куба. Для того чтобы наблюдатель воспринял трехмерную форму объекта на основании только одного вида, надо, чтобы были видны несколько граней этого объекта. Для простых объектов, подобных кубу, должны быть видны как минимум три грани. Вид с несколькими гранями можно получить из одноточечной перспективной проекции с фиксированным центром и с плоскостью проецирования, перпендикулярной направлению взгляда, если предварительно выполнен перенос и/или поворот объекта. Тогда получается реалистичный вид, если только центр проекции не находится слишком близко к объекту.

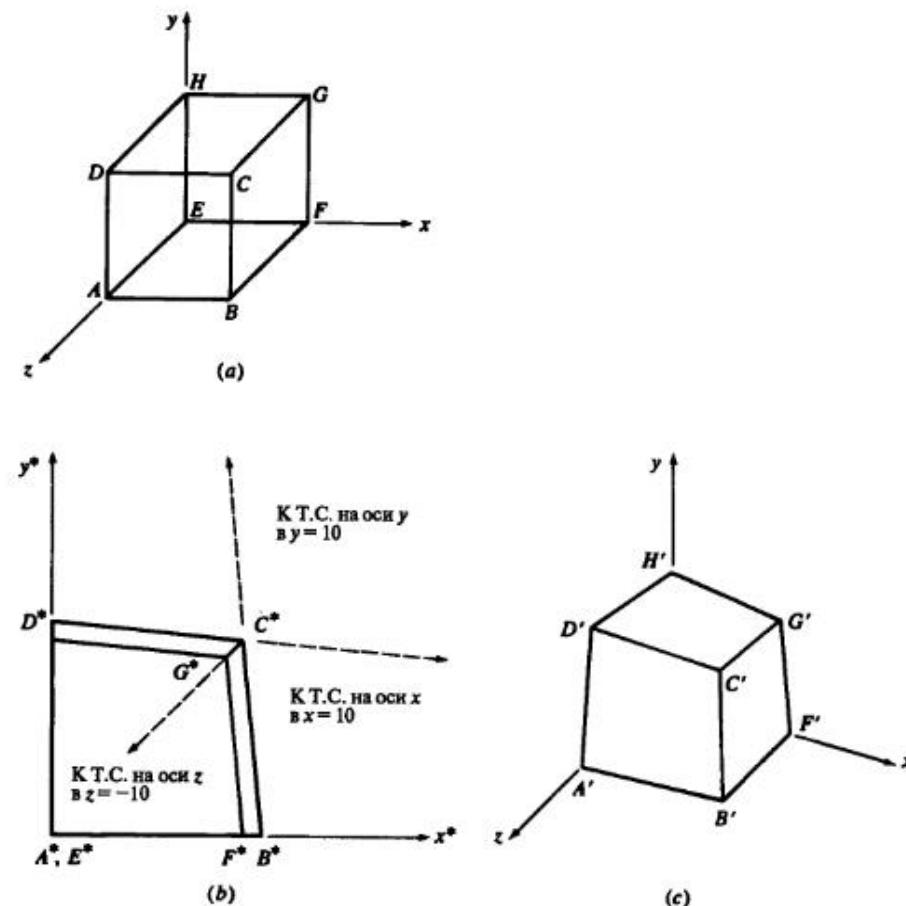
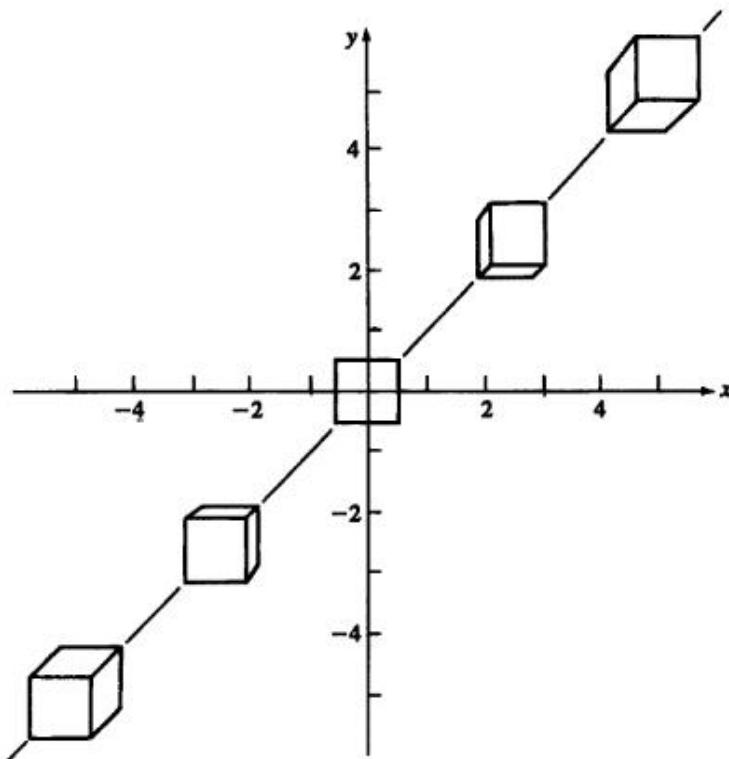


Рис. 3-31 Трехточечная перспектива. (а) Исходный куб; (б) перспективная проекция на плоскость $z = 0$; (с) искаженный куб.

Для начала рассмотрим простой перенос объекта с последующим одноточечным проецированием на плоскость $z = 0$ и с центром проекции в точке $z = z_c$. Требуемое преобразование записывается в виде

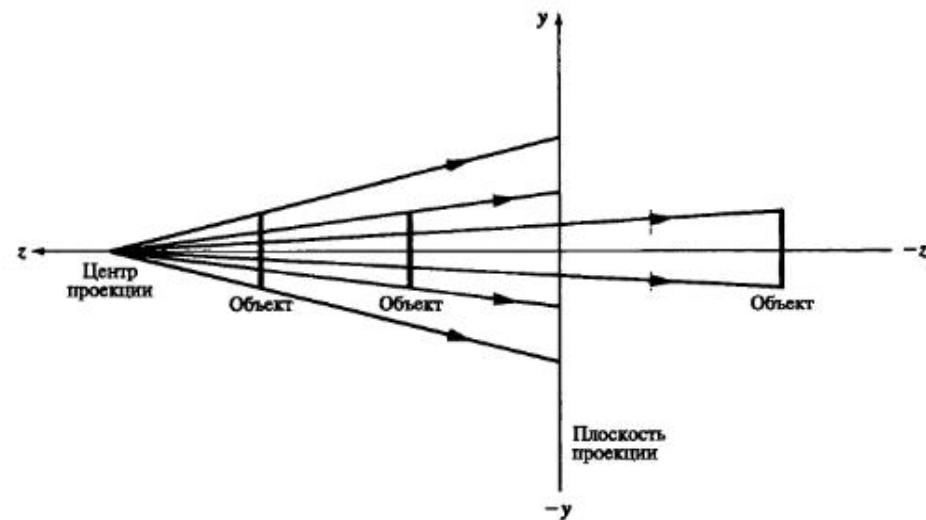
$$\begin{aligned} [T] = [Tr_{xyz}] [Pr_z] &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ l & m & n & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & r \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & r \\ l & m & n & 1 + rn \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1/z_c \\ l & m & n & 1 - n/z_c \end{bmatrix}, \end{aligned} \quad (3-59)$$

где $r = -1/z_c$.

Рис. 3-32 Одноточечная перспективная проекция с переносом в x , y направлениях.

Уравнение (3-59) вместе с рис. 3-32 показывает, что перенос в направлениях x и y открывает дополнительные грани объекта. Перенос в обоих этих направлениях необходим, чтобы открыть три грани простого кубообразного объекта. На рис. 3-32 показаны результаты переноса вдоль прямой $y = x$ отцентрированного относительно начала координат куба и одноточечного проецирования на плоскость $z = 0$. Заметим, что для передней грани показываются истинные размеры и форма.

Уравнение (3-59) также показывает, что перенос вдоль оси z , т. е. к центру проекции или от него, приводит к явному изменению масштаба (из-за элемента $1 - n/z_c$). Этот эффект соответствует физической реальности, так как объекты, находящиеся дальше от наблюдателя, выглядят более мелкими. Заметим, что при приближении центра проекции к бесконечности явление масштабирования исчезает. Этот эффект схематично показан на рис. 3-33. Как изображено на этом рисунке, объект может находиться с любой стороны от центра проекции. Если объект и плоскость проекции находятся по одну сторону от центра, то, как показано на рис. 3-33, получается прямое изображение. Если же объект и плоскость проекции лежат по разные стороны от центра, то получается перевернутое изображение.

Рис. 3-33 Эффект масштабирования при перемещениях вдоль оси z для Одноточечной перспективной проекции.

На рис. 3-34 показаны результаты перемещения объекта во всех трех направлениях. Здесь куб перемещается вдоль трехмерной прямой от $-x = -y = -z$ к $x = y = z$. Заметно очевидное увеличение размера, а также на всех видах заметно сохранение истинной формы, но не размера передней грани.

Эти идеи более подробно изложены в примере.

Пример 3-22 Одноточечная перспективная проекция с переносом

Рассмотрим отцентрированный относительно начала координат единичный куб со следующими координатными векторами

$$[X] = \begin{bmatrix} -0.5 & -0.5 & 0.5 & 1 \\ 0.5 & -0.5 & 0.5 & 1 \\ 0.5 & 0.5 & 0.5 & 1 \\ -0.5 & 0.5 & 0.5 & 1 \\ -0.5 & -0.5 & -0.5 & 1 \\ 0.5 & -0.5 & -0.5 & 1 \\ 0.5 & 0.5 & -0.5 & 1 \\ -0.5 & 0.5 & -0.5 & 1 \end{bmatrix}.$$

Переместим куб на 5 единиц в направлениях x и y и построим перспективную проекцию на плоскость $z = 0$ с центром проекции в $z = z_c = 10$.

Из уравнения (3-59) получаем общую матрицу преобразования

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -0.1 \\ 5 & 5 & 0 & 1 \end{bmatrix}.$$

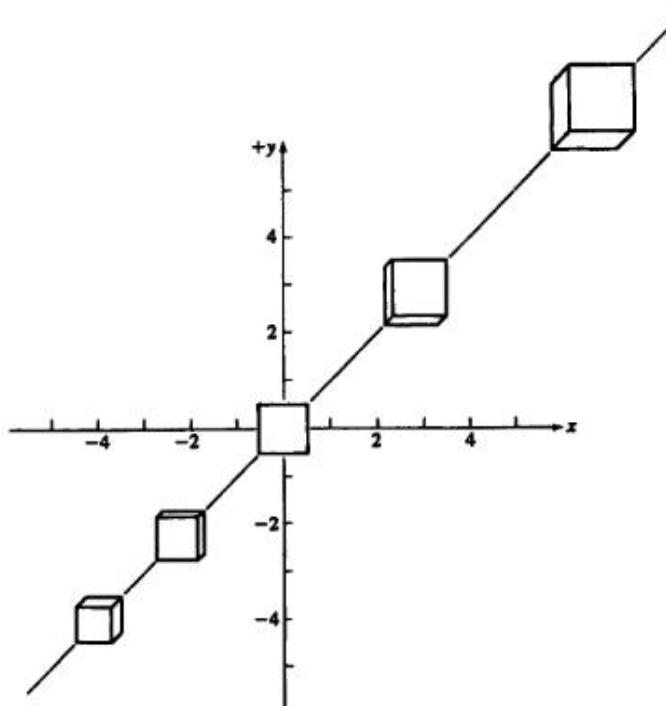


Рис. 3-34 Одноточечная перспективная проекция, объединенная с переносами в x , y , z направлениях.

Преобразованные координаты

$$[X^*] = [X][T] = \begin{bmatrix} 4.5 & 4.5 & 0 & 0.95 \\ 5.5 & 4.5 & 0 & 0.95 \\ 5.5 & 5.5 & 0 & 0.95 \\ 4.5 & 5.5 & 0 & 0.95 \\ 4.5 & 4.5 & 0 & 1.05 \\ 5.5 & 4.5 & 0 & 1.05 \\ 5.5 & 5.5 & 0 & 1.05 \\ 4.5 & 5.5 & 0 & 1.05 \end{bmatrix} = \begin{bmatrix} 4.737 & 4.737 & 0 & 1 \\ 5.789 & 4.737 & 0 & 1 \\ 5.789 & 5.789 & 0 & 1 \\ 4.737 & 5.789 & 0 & 1 \\ 4.286 & 4.286 & 0 & 1 \\ 5.238 & 4.286 & 0 & 1 \\ 5.238 & 5.238 & 0 & 1 \\ 4.286 & 5.238 & 0 & 1 \end{bmatrix}$$

Правый верхний объект на рис. 3-32 изображает этот результат.

Если исходный объект был перемещен на 5 единиц в направлениях x , y , z и была построена одноточечная перспективная проекция на плоскость $z = 0$ с центром проекции в $z = z_c = 20$, тогда из (3-59) следует, что общая матрица преобразования записывается в виде

$$[T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.05 \\ 5 & 5 & 0 & 0.75 \end{bmatrix}.$$

Отметим общее масштабирование, задаваемое значением 0.75 в правом нижнем элементе матрицы преобразования.

Преобразованные координаты равны

$$[X^*] = [X][T] = \begin{bmatrix} 4.5 & 4.5 & 0 & 0.725 \\ 5.5 & 4.5 & 0 & 0.725 \\ 5.5 & 5.5 & 0 & 0.725 \\ 4.5 & 5.5 & 0 & 0.725 \\ 4.5 & 4.5 & 0 & 0.775 \\ 5.5 & 4.5 & 0 & 0.775 \\ 5.5 & 5.5 & 0 & 0.775 \\ 4.5 & 5.5 & 0 & 0.775 \end{bmatrix} = \begin{bmatrix} 6.207 & 6.207 & 0 & 1 \\ 7.586 & 6.207 & 0 & 1 \\ 7.586 & 7.586 & 0 & 1 \\ 6.207 & 7.586 & 0 & 1 \\ 5.806 & 5.806 & 0 & 1 \\ 7.097 & 5.806 & 0 & 1 \\ 7.097 & 7.097 & 0 & 1 \\ 5.806 & 7.097 & 0 & 1 \end{bmatrix}.$$

Результат показан в виде верхнего правого объекта на рис. 3-34.

Несколько граней также будет видно, если использовать вращение объекта. Один поворот откроет по крайней мере две грани объекта, тогда как два и более поворотов вокруг разных осей откроют, как минимум, три грани.

Матрица преобразования для поворота вокруг оси y на угол ϕ и последующего одноточечного перспективного проецирования на плоскость $z = 0$ с центром проекции в $z = z_c$:

$$\begin{aligned} [T] = [R_y][P_{rz}] &= \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/z_c \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \cos \phi & 0 & 0 & \frac{\sin \phi}{z_c} \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & 0 & -\frac{\cos \phi}{z_c} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \quad (3-60)$$

Аналогичным образом матрица преобразования для поворота вокруг оси x на угол θ и последующего одноточечного перспективного проецирования на плоскость $z = 0$ с центром проекции в точке $z = z_c$ имеет вид:

$$\begin{aligned} [T] = [R_x][P_{rz}] &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/z_c \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & 0 & -\frac{\sin \theta}{z_c} \\ 0 & -\sin \theta & 0 & -\frac{\cos \theta}{z_c} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned} \quad (3-61)$$

В обоих уравнениях (3-60) и (3-61) не равны нулю два отвечающих за перспективное преобразование (перспективных) элемента в четвертом столбце матрицы преобразования. Таким образом, один поворот вокруг главной оси, перпендикулярной той оси, на которой лежит центр проекции, эквивалентен двуточечному

перспективному преобразованию. При повороте вокруг оси, на которой лежит центр проекции, такого эффекта нет. Заметим, что для одного поворота перспективный элемент для оси вращения остается неизменным, например, в уравнениях (3-60) и (3-61) элементы q и r соответственно равны нулю.

В общем случае вращение вокруг главной оси не открывает необходимого для адекватного трехмерного представления числа граней — как минимум, трех. Для этого оно должно быть скомбинировано с перемещением вдоль оси. Следующий пример иллюстрирует это.

Пример 3-23 Двухточечное перспективное проецирование с использованием поворота вокруг одной главной оси

Рассмотрим проекцию на плоскость $z = 0$ с центром в точке $z = z_c = 2.5$ куба с рис. 3-35 а, повернутого вокруг оси y на угол $\phi = 60^\circ$, чтобы открылась левая грань, и перемещенного на -2 единицы вдоль y , чтобы открылась верхняя грань.

Используя уравнение (3-38) с $\phi = 60^\circ$, уравнение (3-47) с $z_c = 2.5$ и уравнение (3-14) с $n = 1 = 0$, $m = -2$, получим

$$[T] = [R_y][Tr][P_{rz}] =$$

$$= \begin{bmatrix} 0.5 & 0 & -0.866 & 0 \\ 0 & 1 & 0 & 0 \\ 0.866 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.4 \\ 0 & 0 & 0 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.5 & 0 & 0 & 0.346 \\ 0 & 1 & 0 & 0 \\ 0.866 & 0 & 0 & -0.2 \\ 0 & -2 & 0 & 1 \end{bmatrix}.$$

Преобразованные координаты равны

$$[X^*] = [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 & 0 & 0.346 \\ 0 & 1 & 0 & 0 \\ 0.866 & 0 & 0 & -0.2 \\ 0 & -2 & 0 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.866 & -2 & 0 & 0.8 \\ 1.366 & -2 & 0 & 1.146 \\ 1.366 & -1 & 0 & 1.146 \\ 0.866 & -1 & 0 & 0.8 \\ 0 & -2 & 0 & 1 \\ 0.5 & -2 & 0 & 1.346 \\ 0.5 & -1 & 0 & 1.346 \\ 0 & -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.083 & -2.5 & 0 & 1 \\ 1.192 & -1.745 & 0 & 1 \\ 1.192 & -0.872 & 0 & 1 \\ 1.083 & -1.25 & 0 & 1 \\ 0 & -2 & 0 & 1 \\ 0.371 & -0.485 & 0 & 1 \\ 0.371 & -0.743 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

Результат показан на рис. 3-35 б. Искажение появляется из-за того, что центр проекции расположен слишком близко к кубу. Отметим схождение параллельных осей x и z прямых линий к точкам схода, лежащим на оси x . Эти точки схода определяются в примере 3-25 из разд. 3-17.

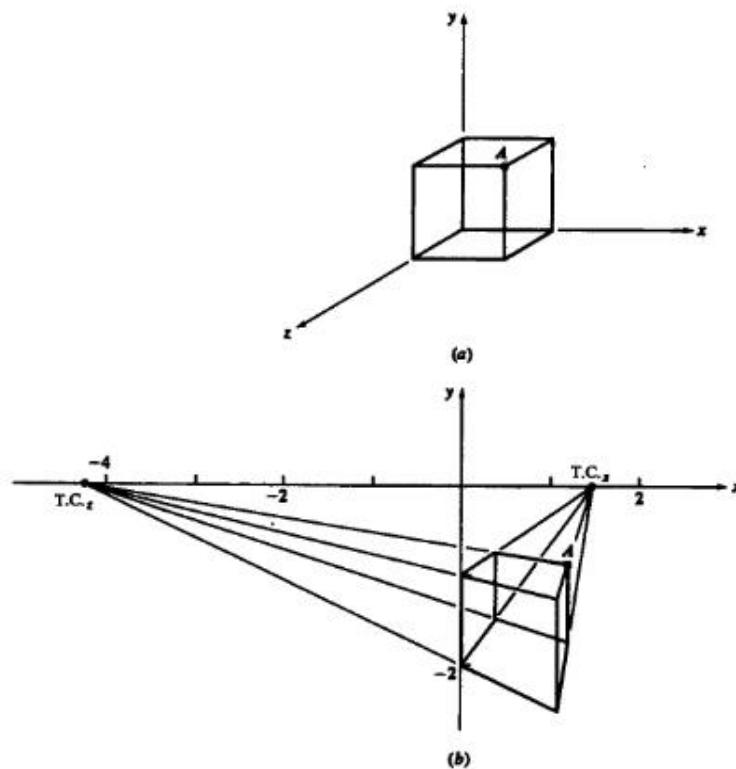


Рис. 3-35 Двухточечная перспективная проекция с поворотом вокруг одной оси.

Аналогичным образом трехточечное перспективное преобразование выполняется с помощью вращения вокруг двух или более главных осей и последующего одноточечного перспективного преобразования. Например, поворот вокруг оси y , потом поворот вокруг оси x и перспективное проецирование на плоскость $z = 0$ с центром проекции в точке $z = z_c$ имеет следующую матрицу преобразования

$$[T] = [R_y][R_x][P_{rz}] =$$

$$= \begin{bmatrix} \cos \phi & 0 & -\sin \phi & 0 \\ 0 & 1 & 0 & 0 \\ \sin \phi & 0 & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1/z_c \\ 0 & 0 & 0 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & \frac{\sin \phi \cos \theta}{z_c} \\ 0 & \cos \theta & 0 & -\frac{\sin \theta}{z_c} \\ \sin \phi & -\cos \phi \sin \theta & 0 & -\frac{\cos \phi \cos \theta}{z_c} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-62)$$

Отметим три ненулевых перспективных элемента. Объект можно также переместить, если перемещение происходит после вращения, тогда результирующая матрица преобразования равна

$$\begin{aligned} [T] &= [R_y][R_x][Tr][P_{rz}] = \\ &= \begin{bmatrix} \cos \phi & \sin \phi \sin \theta & 0 & \frac{\sin \phi \cos \theta}{z_c} \\ 0 & \cos \theta & 0 & -\frac{z_c}{\sin \theta} \\ \sin \phi & -\cos \phi \sin \theta & 0 & -\frac{z_c \cos \phi \cos \theta}{z_c} \\ l & m & 0 & 1 - \frac{z_n}{z_c} \end{bmatrix}. \end{aligned} \quad (3-63)$$

Отметим здесь очевидный масштабирующий эффект перемещения вдоль z . Результаты преобразования будут различными, если поменять порядок выполнения поворотов или перенос выполнять до вращения.

Пример 3-24 Трехточечная перспективная проекция с поворотом вокруг двух осей

Рассмотрим куб на рис. 3-35 а, повернутый вокруг оси y на угол $\phi = -30^\circ$, вокруг оси x на $\theta = 45^\circ$ и спроектированный на плоскость $z = 0$ с центром в точке $z = z_c = 2.5$.

Из уравнения (3-62)

$$[T] = \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразованные координаты равны

$$\begin{aligned} [X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} -0.5 & -0.612 & 0 & 0.755 \\ 0.366 & -0.966 & 0 & 0.614 \\ 0.366 & -0.259 & 0 & 0.331 \\ -0.5 & 0.095 & 0 & 0.472 \\ 0 & 0 & 0 & 1 \\ 0.866 & -0.354 & 0 & 0.859 \\ 0.866 & 0.354 & 0 & 0.576 \\ 0 & 0.707 & 0 & 0.717 \end{bmatrix} = \begin{bmatrix} -0.662 & -0.811 & 0 & 1 \\ 0.596 & -1.574 & 0 & 1 \\ 1.107 & -0.782 & 0 & 1 \\ -1.059 & 0.201 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1.009 & -0.412 & 0 & 1 \\ 1.504 & 0.614 & 0 & 1 \\ 0 & 0.986 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Результат изображен на рис. 3-36.

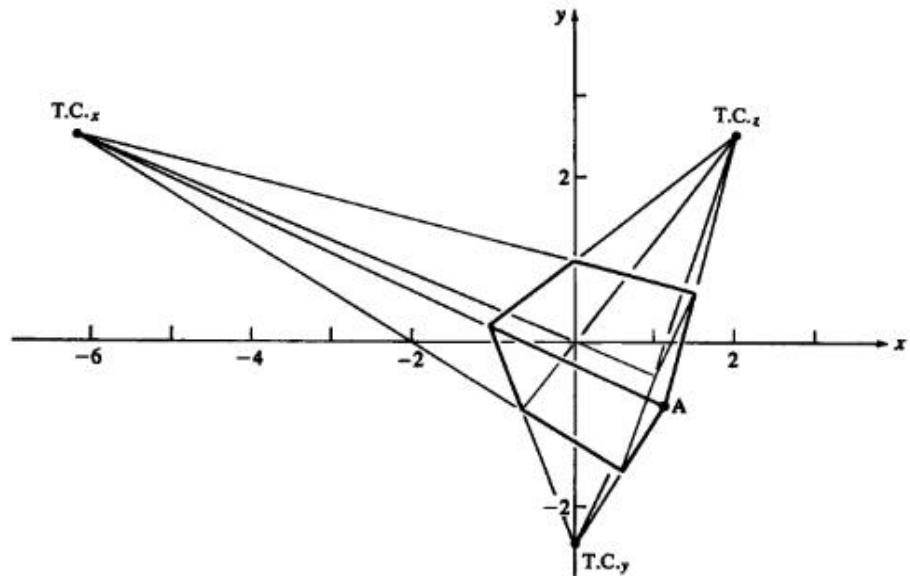


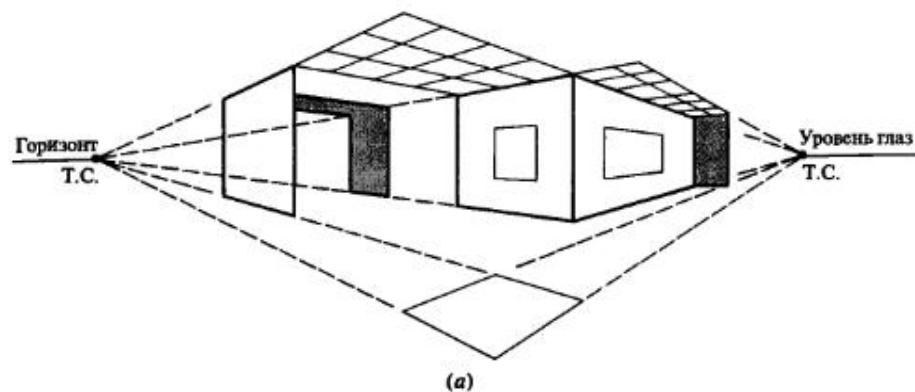
Рис. 3-36 Трехточечная перспективная проекция с поворотом вокруг двух осей.

Из этих результатов становится ясно, что одно-, дву- или трехточечное перспективное преобразование можно сконструировать с помощью поворотов и переносов вокруг и вдоль главных осей с последующим одноточечным перспективным преобразованием с центром проекции, расположенным на одной из главных осей. Эти результаты также справедливы для поворота вокруг произвольной оси в пространстве. Следовательно, при использовании в графической системе парадигмы с фиксированным центром проекции и манипулируемым объектом, необходимо обеспечить только построение одноточечной перспективной проекции на плоскость $z = 0$ с центром проекции на оси z .

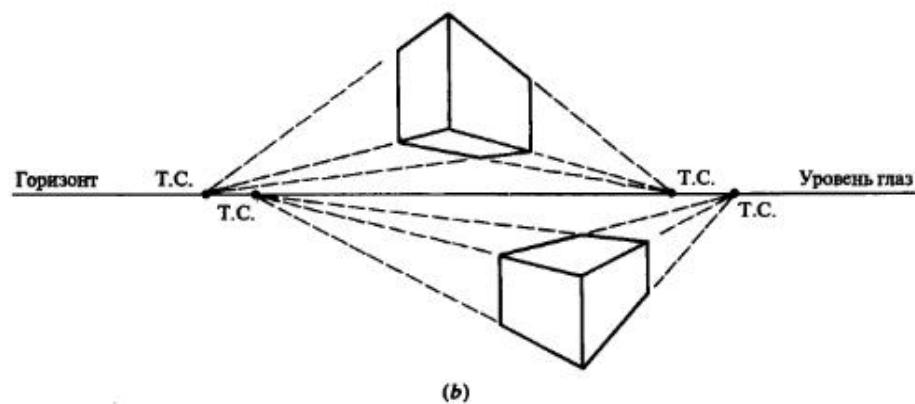
3-17 ТОЧКИ СХОДА

Как показано на рис. 3-37 а при построении перспективного вида объекта используется линия горизонта, расположенная обычно на уровне глаз. Главные точки схода — это те точки на линии горизонта, в которые сходятся прямые, параллельные в исходном пространстве главным осям. В общем случае различные множества параллельных прямых имеют различные главные точки схода, как это показано на рис. 3-37 б. Для плоскостей объекта, расположенных наклонно относительно исходных главных осей, точки схода лежат выше или ниже линии горизонта. Как показано на рис. 3-37 с, такие точки часто называются следом точек.

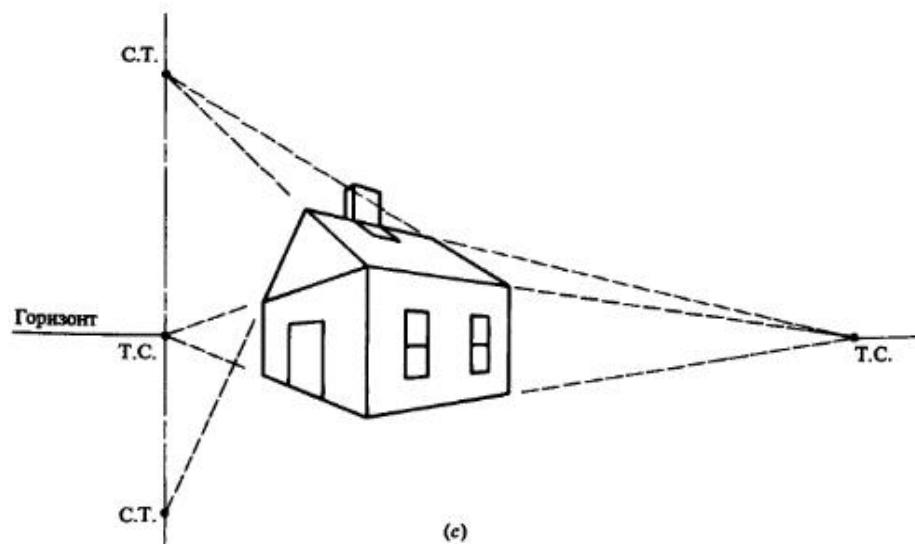
Представляют интерес два метода определения точек схода. В первом просто вычисляется точка пересечения пары преобразованных спроектированных параллельных прямых. Второй метод сложен, но зато дает более точные результаты. В



(a)



(b)



(c)

Рис. 3-37 Следы точек и точки схода.

в этом методе в нужную позицию и ориентацию преобразуется объект, стороны которого параллельны исходным главным осям. Потом применяется одноточечное перспективное проецирование. Затем результирующая матрица преобразования (см. уравнение 3-63) применяется к точкам, расположенным в бесконечности на главных осях. Полученные в результате этого обычные координаты есть главные точки схода для этого объекта. Для нахождения следов точек, возникающих для наклонных плоскостей, сначала находят расположенные в бесконечности точки на наклонной плоскости и затем подвергают их преобразованию.

Эти методы проиллюстрированы на нескольких примерах. В первом из них для определения точек схода используется пересечение преобразованных прямых.

Пример 3-25 Главные точки схода, определяемые по пересечению прямых

Из примера 3-23 преобразованные координатные векторы для пары отрезков, один из которых проходит через точку A (см. рис. 3-35 а), и первоначально параллельных соответственно осям x и z , равны

$$\begin{array}{ll} \textcircled{4} & \begin{bmatrix} 1.083 & -1.25 & 0 & 1 \end{bmatrix} \\ \textcircled{3} & \begin{bmatrix} 1.192 & -0.872 & 0 & 1 \end{bmatrix} \\ \textcircled{8} & \begin{bmatrix} 0 & -1 & 0 & 1 \end{bmatrix} \\ \textcircled{7} & \begin{bmatrix} 0.371 & 0.743 & 0 & 1 \end{bmatrix} \end{array} \quad \text{и} \quad \begin{array}{ll} \textcircled{3} & \begin{bmatrix} 1.192 & -0.872 & 0 & 1 \end{bmatrix} \\ \textcircled{7} & \begin{bmatrix} 0.371 & -0.743 & 0 & 1 \end{bmatrix} \\ \textcircled{4} & \begin{bmatrix} 1.083 & -1.25 & 0 & 1 \end{bmatrix} \\ \textcircled{8} & \begin{bmatrix} 0 & -1 & 0 & 1 \end{bmatrix} \end{array}$$

Здесь цифры в кружках относятся к строкам исходной и преобразованной матрицы из примера 3-23. Уравнения пары прямых, параллельных в исходном пространстве оси x таковы:

$$\begin{aligned} y &= 3.468x - 5.006, \\ y &= 0.693x - 1. \end{aligned}$$

Решение этой системы дает точку схода $[VP_z] = [1.444 \ 0]$.

Уравнения пары прямых, параллельных в исходном пространстве оси z , имеют вид:

$$\begin{aligned} y &= -0.157x - 0.685 \\ y &= -0.231x - 1. \end{aligned}$$

Решение дает точку схода $[VP_z] = [-4.333 \ 0]$.

Эти точки показаны на рис. 3-35 б.

Во втором примере для нахождения точек схода используется преобразование расположенных в бесконечности точек на главных осях.

Пример 3-26 Главные точки схода, найденные с помощью преобразования

Общая матрица преобразования из примера 3-24

$$[T] = \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Преобразование расположенных в бесконечности на осях x , y и z точек дает

$$[VP][T] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ = \begin{bmatrix} -6.142 & 2.5 & 0 & 1 \\ 0 & -2.5 & 0 & 1 \\ 2.04 & 2.5 & 0 & 1 \end{bmatrix}.$$

Эти точки схода показаны на рис. 3-36.

В третьем примере для нахождения следов точек используется преобразование точек, расположенных в бесконечности на наклонных плоскостях.

Пример 3-27 Следы точек, полученные с помощью преобразования

Рассмотрим простую треугольную призму на рис. 3-38 а. Координатные векторы призмы

$$[X] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0.5 & 0.5 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0.5 & 0.5 & 0 & 1 \end{bmatrix}.$$

Применяя общее преобразование из примера 3-24, получим преобразованные координаты

$$[X^*] = [X][T] = \\ = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0.5 & 0.5 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0.5 & 0.5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ = \begin{bmatrix} -0.5 & -0.612 & 0 & 0.755 \\ 0.366 & -0.966 & 0 & 0.614 \\ -0.067 & -0.436 & 0 & 0.543 \\ 0 & 0 & 0 & 1 \\ 0.866 & -0.354 & 0 & 0.859 \\ 0.433 & 0.177 & 0 & 0.788 \end{bmatrix} = \\ = \begin{bmatrix} -0.662 & -0.811 & 0 & 1 \\ 0.596 & -1.574 & 0 & 1 \\ -0.123 & -0.802 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1.009 & -0.412 & 0 & 1 \\ 0.55 & 0.224 & 0 & 1 \end{bmatrix}.$$

На рис. 3-38 б изображена преобразованная призма.

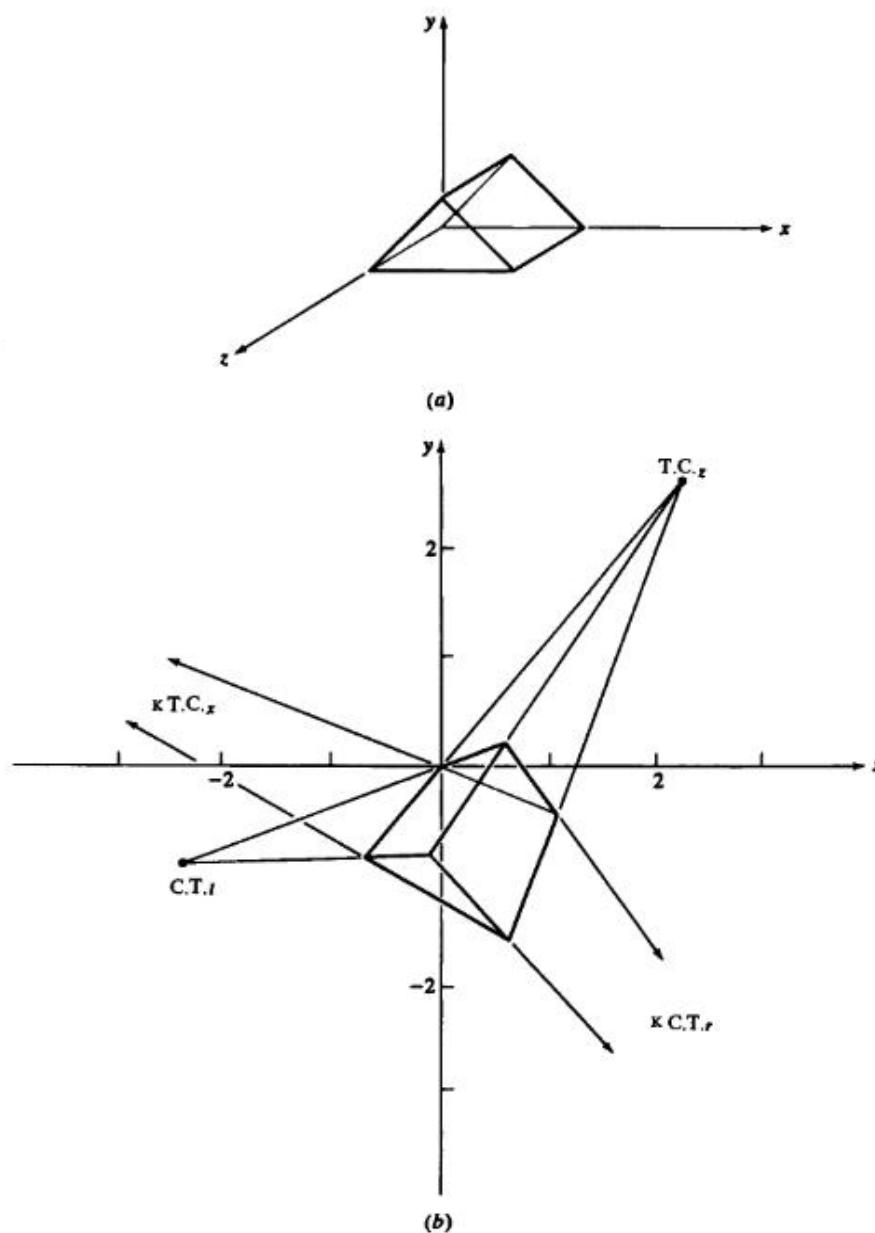


Рис. 3-38 Следы точек.

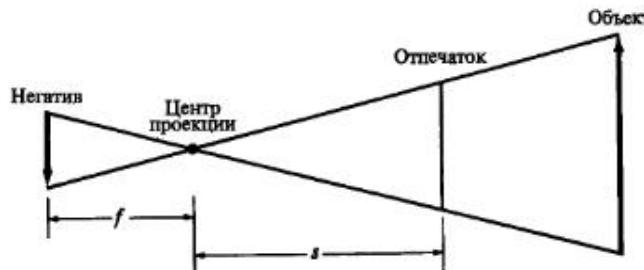


Рис. 3-39 Фотография как перспективная проекция.

Направляющие косинусы наклонных ребер левой верхней плоскости призмы до преобразования равны $[0.5 \ 0.5 \ 0]$. Таким образом, точка, лежащая в бесконечности в этом направлении, имеет однородные координаты $[1 \ 1 \ 0 \ 0]$.

Аналогично $[-0.5 \ 0.5 \ 0]$ — направляющие косинусы наклонных ребер правой верхней плоскости призмы до преобразования. Таким образом, точка, лежащая в бесконечности в этом направлении, имеет координаты $[-1 \ 1 \ 0 \ 0]$.

Применяя преобразование к только что полученным точкам и к точкам, лежащим в бесконечности на главных осях, получим

$$\begin{aligned} [VP][T] &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0.866 & -0.354 & 0 & -0.141 \\ 0 & 0.707 & 0 & -0.283 \\ -0.5 & -0.612 & 0 & -0.245 \\ 0.866 & 0.354 & 0 & -0.424 \\ -0.866 & 1.061 & 0 & -0.141 \end{bmatrix} = \\ &= \begin{bmatrix} -6.142 & 2.5 & 0 & 1 \\ 0 & -2.5 & 0 & 1 \\ 2.041 & 2.5 & 0 & 1 \\ -2.041 & -0.833 & 0 & 1 \\ 6.142 & -7.5 & 0 & 1 \end{bmatrix} VP_z \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1/f \\ 0 & 0 & 0 & 1 \end{bmatrix} TP_z. \end{aligned}$$

Точки схода и следы точек изображены на рисунке 3-38 б. Как мы и ожидали, точки VP_x , VP_y и VP_z совпадают с полученными в примере 3-26.

3-18 ФОТОГРАФИЯ И ПЕРСПЕКТИВНОЕ ПРЕОБРАЗОВАНИЕ

Фотография — это перспективная проекция. Схема работы камеры в общем виде изображена на рис. 3-39. Центр проекции — это точка фокуса линзы камеры. Удобно рассматривать создание негатива и отпечатка с него как два разных процесса.

На рис. 3-40 а показана схема создания исходного негатива. Для удобства поместим негатив в плоскость $z = 0$ так, чтобы центр проекции и сцена находились

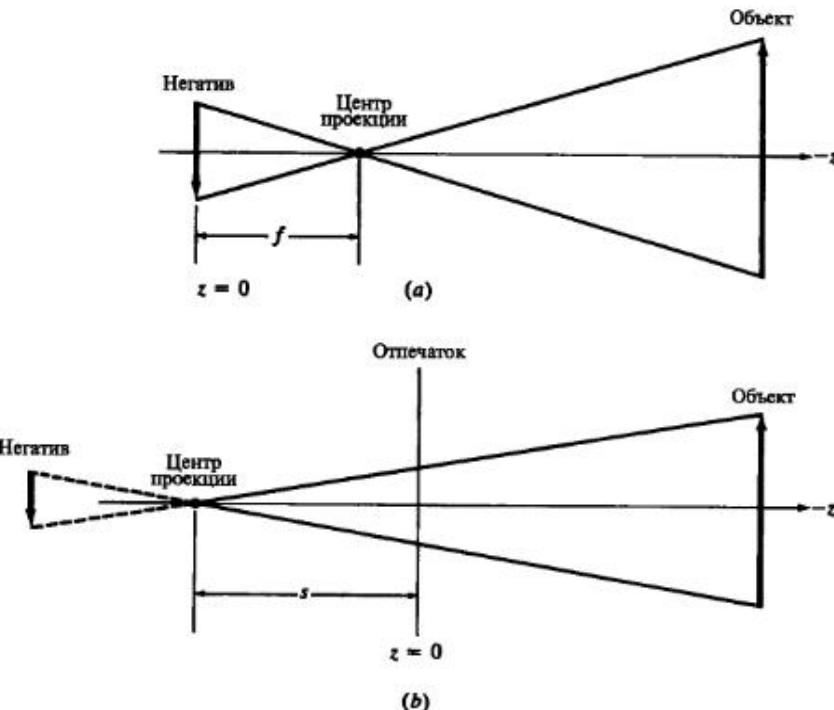


Рис. 3-40 Схема фотографической перспективы. (а) Создание исходного негатива; (б) создание оттиска.

в отрицательном полупространстве $z < 0$. Перспективное проецирование на плоскость $z = 0$ (негатив) порождает следующее преобразование:

$$[T_n] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/f \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (3-64)$$

где f — это фокусное расстояние линзы. Заметим, что на негативе формируется перевернутое изображение объекта. Конкретнее,

$$[x \ y \ z \ 1] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/f \\ 0 & 0 & 0 & 1 \end{bmatrix} = [x \ y \ 0 \ 1 + z/f]$$

и

$$x^* = \frac{x f}{f + z}, \quad y^* = \frac{y f}{f + z}.$$

Здесь для $f + z < 0$ значения x^* и y^* имеют противоположный знак по сравнению с x и y , следовательно, на негативе образуется перевернутое изображение.

На рисунке 3-40 *b* изображена схема создания отпечатка с фотографического негатива. Здесь s обозначает расстояние от фокуса линзы увеличителя до бумаги. Предполагается, что бумага находится в $z = 0$. Тогда преобразование перспективного проецирования имеет вид:

$$[T_p] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/s \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3-65)$$

Заметим, что на отпечатке формируется прямое изображение. А именно,

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/s \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 - \frac{z}{s} \end{bmatrix}$$

и

$$x^* = \frac{xs}{s-z}, \quad y^* = \frac{ys}{s-z}.$$

Для объекта значение $z < 0$, тогда $s - z > 0$ и координаты x^* , y^* имеют тот же знак, что и x , y , т. е. формируется прямое изображение.

3-19 СТЕРЕОГРАФИЧЕСКАЯ ПРОЕКЦИЯ

Во многих приложениях важно увеличение восприятия трехмерной глубины сцены. Для восприятия глубины у системы глаз–мозг существует два основных типа подсказок: монокулярные и бинокулярные, в зависимости от того, сколько глаз используется – один или два. Основными монокулярными подсказками являются:

- перспектива – схождение в одну точку параллельных прямых;
- параллакс движения – при горизонтальном движении головы кажется, что ближние объекты сдвигаются на плоскости проекции больше, чем дальние;
- относительный размер знакомых наблюдателю объектов;
- перекрытие – более близкий объект загораживает остальные и располагается перед более отдаленным;
- свет и тени;
- ослабление атмосферой и неспособность глаза из-за этого различить мелкие детали удаленных объектов;
- фокусирующая аккомодация – для объектов, расположенных на разных расстояниях, требуются разные усилия фокусирующих мускулов глаза;

Основными бинокулярными подсказками являются:

- углы сближения оптических осей глаз;

- несогласованность на сетчатке – разница в местоположении спроектированных на сетчатку глаза объектов интерпретируется как различие в расстоянии до глаз.

Монокулярное зрение дает лишь слабое восприятие трехмерной глубины. Однако бинокулярное зрение дает очень сильное восприятие глубины, потому что система глаз–мозг смешивает в одно два различных изображения, созданных для каждого глаза. Стереография пытается создать изображение с характеристиками, аналогичными настоящему бинокулярному зрению. Для генерации стереоизображений существуют несколько методов (см. [3-4] и [3-5]). Все они основаны на том, что для левого и правого глаза готовятся разные изображения.

В двух методах, называемых цветовым анаглифом и поляризованным анаглифом методами, для получения левым и правым глазами корректных и различных изображений используются фильтры. Если рассказать в нескольких словах, то в первом случае двумя разными цветами создаются два изображения – одно для левого и другое для правого глаза. При просмотре через соответствующие фильтры левый глаз видит левое изображение, а правый – только правое. Система глаз–мозг объединяет оба двумерных изображения в одно трехмерное с правильными цветами. Во втором случае вместо цветовых фильтров используются поляризованные.

В третьем методе используется попеременное проецирование видов для левого и правого глаза. Соответствующее устройство просмотра синхронизировано так, чтобы свет не попадал в другой глаз.

В четвертом методе, автостереоскопии, никакого специального оборудования для просмотра не требуется. Метод основан на использовании плоских или линзообразных экранов. Изображения называются параллаксными стереограммами, параллаксными панорамаграммами и панорамными параллаксными стереограммами. Более полную информацию по этому вопросу можно найти в [3-5].

Во всех этих методах требуется проецирование объекта на плоскость из двух различных центров проекции – одно для правого глаза и одно для левого. На рис. 3-41 показано проецирование на плоскость $z = 0$ точки P из центров проекции в E_L ($-e, 0, d_e$) и E_R ($e, 0, d_e$), соответствующих левому и правому глазу.

Как показано на рис. 3-41 *b*, для удобства центр проекции для левого глаза перенесен таким образом, чтобы он лежал на оси z . Используя подобные треугольники, получим

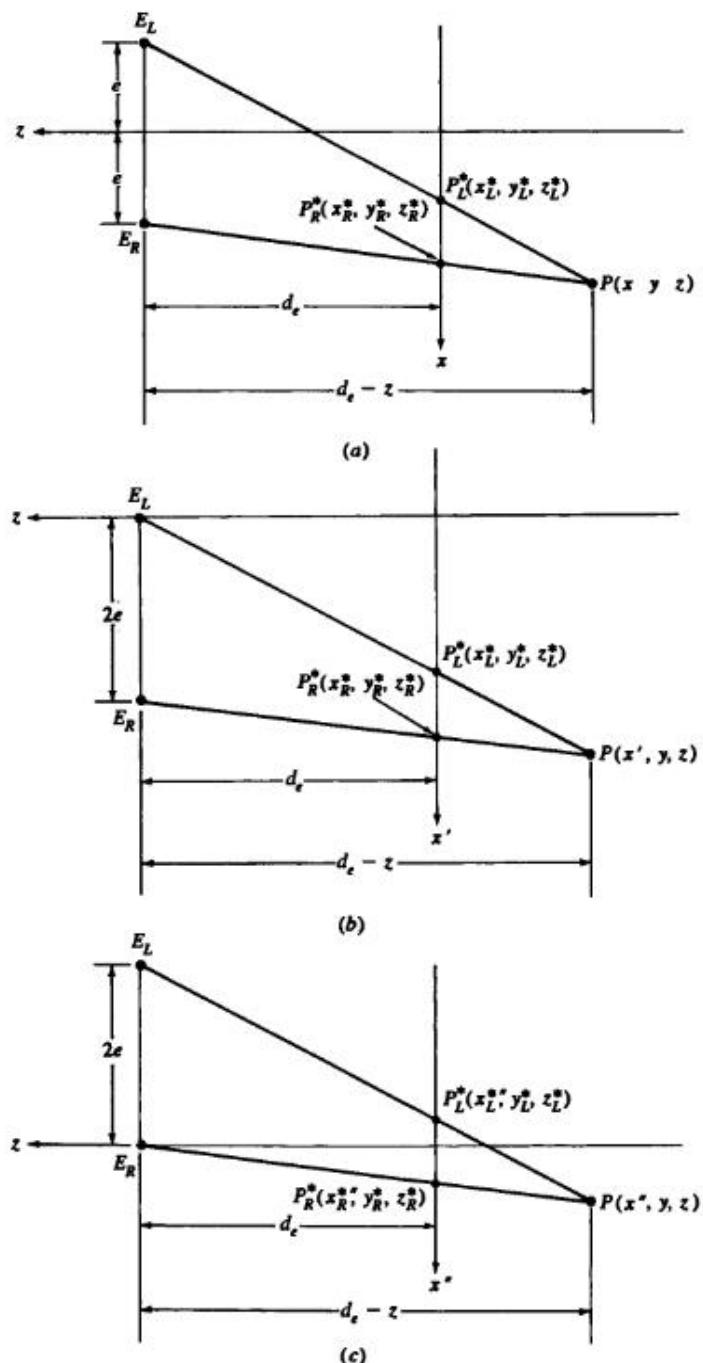
$$\frac{x_L^{*'}}{d_e} = \frac{x'}{d_e - z}$$

и

$$x_L^{*'} = \frac{x'}{1 - z/d_e} = \frac{x'}{1 + rz},$$

где

$$r = -1/d_e.$$

Рис. 3-41 Стереографическая проекция на плоскость $z = 0$.

Аналогичным образом перенеся центр проекции для правого глаза на ось z (как показано на рис. 3-41 с) и снова используя подобные треугольники, получим

$$\frac{x_R^{**}}{d_e} = \frac{x''}{d_e - z}$$

и

$$x_R^{**} = \frac{x''}{1 - z/d_e} = \frac{x''}{1 + rz}.$$

Так как для каждого глаза значение $y = 0$, то оба спроектированных значения y имеют вид:

$$y^* = \frac{y}{1 - z/d_e} = \frac{y}{1 + rz}.$$

Эквивалентные матрицы преобразования 4×4 построения видов для правого и левого глаза следующие:

$$\begin{aligned} [S_L] &= [Tr_z][P_{rz}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ e & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/d_e \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/d_e \\ e & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (3-66)$$

и

$$[S_R] = [Tr_z][P_{rz}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/d_e \\ e & 0 & 0 & 1 \end{bmatrix}. \quad (3-67)$$

Следовательно, стереографическая проекция получается при преобразовании сцены с помощью уравнений (3-66), (3-67) и вывода обоих изображений.

Существует ряд методов для вывода изображений стереографических проекций. Многие люди смогут создать стереоизображение без помощи каких-либо технических средств. Один из методов, требующий некоторой тренировки, состоит в следующем: сначала сфокусировать глаза в бесконечности, затем, не изменяя фокуса, постепенно передвигая стереопары и держа их примерно на расстоянии локтя, добиться стереоэффекта.

Бинокулярное смещение стереопар улучшается при использовании небольшой непрозрачной маски, например полоски черного картона шириной приблизительно в дюйм. Как показано на рис. 3-42, маску помещают между глазами и стереопарой и передвигают вперед и назад до получения нужного положения. В этом

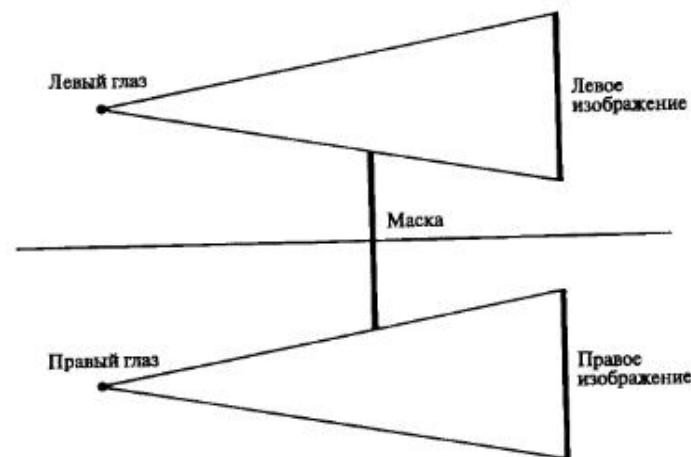


Рис. 3-42 Простой метод просмотра стереопар.

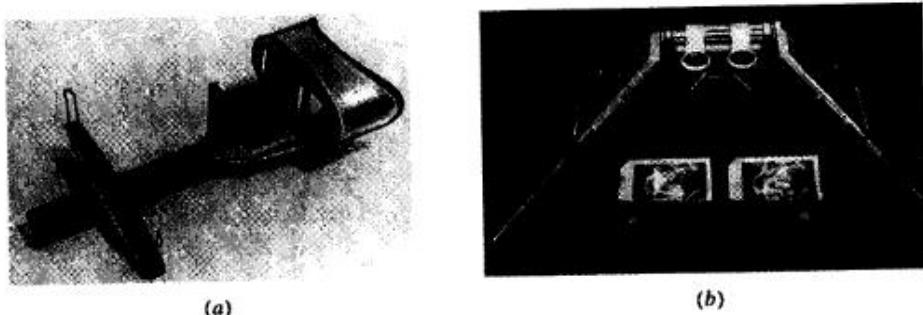


Рис. 3-43 Стереоскопы. (а) Стереоптикон Брустера; (б) типичный лабораторный инструмент.

положении левый глаз видит только левое изображение стереопары, а правый глаз — правое.

Два более специализированных устройства для просмотра стереопар изображены на рис. 3-43 *a* и 3-43 *b*. На рис. 3-43 *a* показан популярный в начале нашего века стереоптикон Брустера, а на рис. 3-43 *b* — типичный современный лабораторный стереоскоп. Оба устройства относятся к простым стереоскопам с фокусирующими линзами.

Хотя уравнения (3-66) и (3-67) задают основное преобразование для генерации стереопар, их надо модифицировать для приспособления к различным схемам стереопросмотровых устройств. Схема простого фокального стереоскопа приведена на рис. 3-44. Стереопары в нем располагаются на фокусном расстоянии f от линз. Стереоизображение реконструируется на расстоянии I от линз. Как было ранее упомянуто, стереоугол γ , очень сильно влияющий на бинокулярное смещение, связан со сближением оптических осей глаз. У обычного человека расстояние



Рис. 3-44 Схема фокусного плоского стереоскопа.

между зрачками глаз приблизительно равно 60 мм. Эксперименты показали ([3-4]), что наилучший стереоэффект наблюдается на расстоянии приблизительно 300 мм. Следовательно, наилучший стереоэффект достигается для $\operatorname{tg}(\gamma/2) = 1/5$. Как показано на рис. 3-44, для достижения бинокулярного смещения стереопары должны быть разнесены на расстояние $2w$ при условии, что они размещены на фокусном расстоянии f .

Обозначив через $2D$ на рис. 3-44 расстояние между линзами стереоскопа, из подобия треугольников получим

$$\frac{D-w}{f} = \frac{D}{I} = \frac{1}{5}.$$

Таким образом, разнесение на расстояние

$$w = D - \frac{f}{5} \quad (3-68)$$

дает наилучший стереоэффект. Для стереопары, просматриваемой через простой линзовый стереоскоп, уравнения (3-66) и (3-67) записываются в виде

$$[S_L] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/f \\ w & 0 & 0 & 1 \end{bmatrix} \quad (3-69a)$$

и

$$[S_R] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/f \\ -w & 0 & 0 & 1 \end{bmatrix}. \quad (3-69b)$$

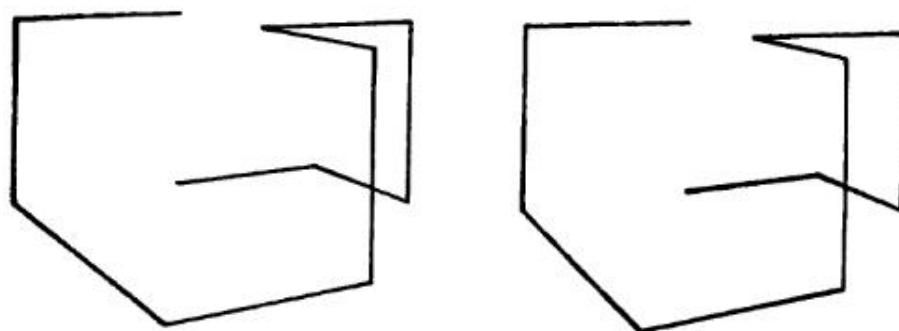


Рис. 3-45 Стереопара.

Типичная стереопара для простого линзового стереоскопа показана на рис. 3-45.

Расстояние между стереопарой, полученное из уравнения (3-68) для типичных значений D и f , довольно мало. Следовательно, можно просматривать только маленькие стереоизображения. Чтобы увеличить размеры изображений для увеличения расстояния между ними используются зеркала или призмы. Типичный зеркальный стереоскоп изображен на рис. 3-43 b. Соответствующая схема приведена на рис. 3-46. Снова используя подобие треугольников, получим расстояние между стереопарой

$$\frac{w_0 - w}{f} = \frac{D}{I} = \frac{1}{5}$$

и

$$w = \frac{w_0 - f}{5}. \quad (3-70)$$

Создание типичной стереопары проиллюстрировано на примере.

Пример 3-28 Создание стереопары

Рассмотрим изображенный на рис. 3-45 простой трехмерный проволочный объект с координатами

$$[X] = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 1 \\ -1 & 0 & 2 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{bmatrix}.$$

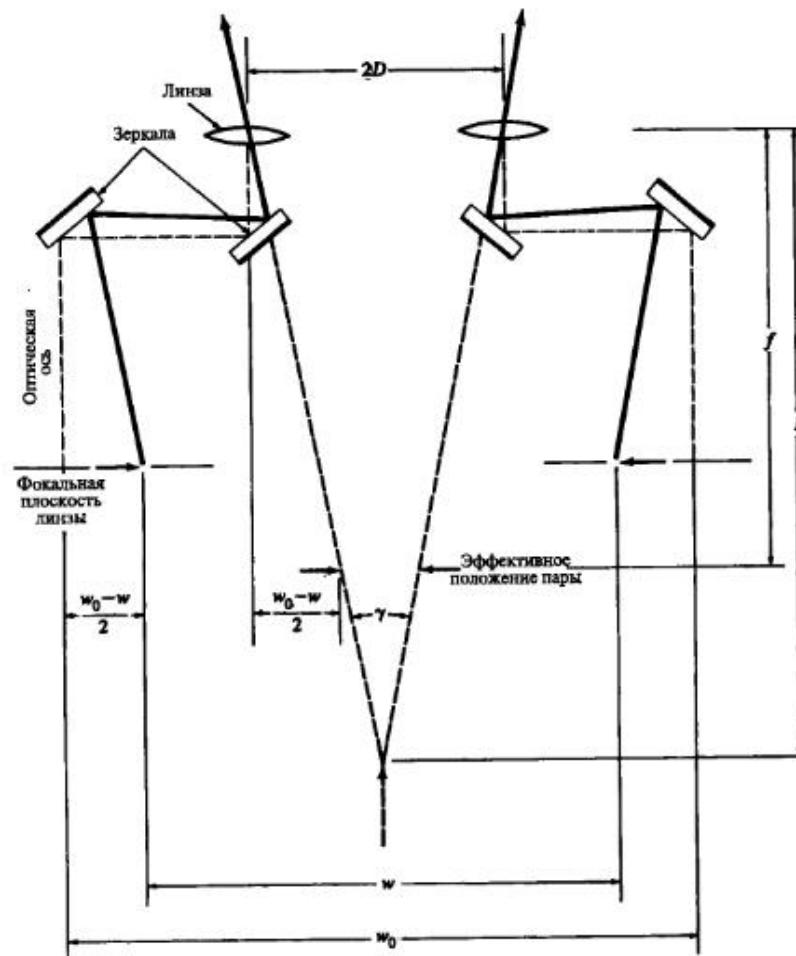


Рис. 3-46 Схема зеркального стереоскопа.

Для визуализации фигуру сначала повернем на 20° вокруг оси y , а затем перенесем на -1.5 единицы в направлении z . Результатирующее преобразование будет

$$[V] = [R_y][T_z] = \begin{bmatrix} 0.94 & 0 & -0.342 & 0 \\ 0 & 1 & 0 & 0 \\ 0.342 & 0 & 0.94 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1.5 & 1 \end{bmatrix} = \\ = \begin{bmatrix} 0.94 & 0 & -0.342 & 0 \\ 0 & 1 & 0 & 0 \\ 0.342 & 0 & 0.94 & 0 \\ 0 & 0 & -1.5 & 1 \end{bmatrix}.$$

Стереопара создается для просматривания с помощью простого фокусирующего плоского стереоскопа. Расстояние между линзами $2D = 4$ дюймам

и фокусное расстояние тоже равно 4 дюймам. Из уравнения (3-68) следует, что расстояние между стереопарой

$$w = 2 - 4/5 = 1.2''.$$

Используя уравнение (3-69), получим преобразования для левого и правого стереоизображений

$$[S_L] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -0.25 \\ 1.2 & 0 & 0 & 1 \end{bmatrix} \quad \text{и} \quad [S_R] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -0.25 \\ -1.2 & 0 & 0 & 1 \end{bmatrix}.$$

Объединение этих преобразований с видовым дает

$$[C_L] = [V][S_L] \quad \text{и} \quad [C_R] = [V][S_R],$$

$$[C_L] = \begin{bmatrix} 0.94 & 0 & 0 & 0.086 \\ 0 & 1 & 0 & 0 \\ 0.342 & 0 & 0 & -0.235 \\ 1.2 & 0 & 0 & 1.38 \end{bmatrix}$$

и

$$[C_R] = \begin{bmatrix} 0.94 & 0 & 0 & 0.086 \\ 0 & 1 & 0 & 0 \\ 0.342 & 0 & 0 & -0.235 \\ -1.2 & 0 & 0 & 1.375 \end{bmatrix}.$$

Преобразованные координаты для левого и правого изображений вычисляются следующим образом:

$$[X_L^*] = [X][C_L] \quad \text{и} \quad [X_R^*] = [X][C_R],$$

$$[X_L^*] = \begin{bmatrix} 0.873 & 0 & 0 & 1 \\ 1.465 & 0 & 0 & 1 \\ 2.025 & 0 & 0 & 1 \\ 2.025 & 0.816 & 0 & 1 \\ 1.353 & 0.877 & 0 & 1 \\ 2.081 & 1.105 & 0 & 1 \\ 2.081 & 0 & 0 & 1 \\ 1.152 & 0 & 0 & 1 \\ 0.202 & 0 & 0 & 1 \\ 0.202 & 0 & 0 & 1 \\ 0.202 & 0.775 & 0 & 1 \end{bmatrix} \quad \text{и} \quad [X_R^*] = \begin{bmatrix} -0.873 & 0 & 0 & 1 \\ -0.178 & 0 & 0 & 1 \\ 0.067 & 0 & 0 & 1 \\ 0.067 & 0.816 & 0 & 1 \\ -0.753 & 0.877 & 0 & 1 \\ -0.57 & 1.105 & 0 & 1 \\ -0.57 & 0 & 0 & 1 \\ -1.776 & 0 & 0 & 1 \\ -1.659 & 0 & 0 & 1 \\ -1.659 & 0 & 0 & 1 \\ -1.659 & 0.775 & 0 & 1 \end{bmatrix}.$$

Результат изображен на рис. 3-45. Заметим, что при создании конечно-го изображения эти данные должны быть промасштабированы для каждого конкретного устройства вывода, с тем чтобы обеспечить корректное физическое значение для w .

3-20 СРАВНЕНИЕ ДВУХ МЕТОДОВ ПОСТРОЕНИЯ ПРОЕКЦИЙ – С ФИКСИРОВАННЫМ ОБЪЕКТОМ И С ФИКСИРОВАННЫМ ЦЕНТРОМ ПРОЕКЦИИ

Метод проецирования, в котором объект зафиксирован, а перемещается центр проекции, может быть легко преобразован в обсуждавшийся ранее метод с перемещающимся объектом и фиксированным центром. Представляет интерес два случая. В первом, более простом, предполагается, что плоскость проекции перпендикулярна направлению взгляда, направленного из центра проекции к сцене. Во втором случае перпендикулярность не предполагается. Здесь будет рассмотрен только первый случай.

Когда плоскость проекции перпендикулярна направлению взгляда, с помощью описанной ниже процедуры построение проекции можно свести к случаю с движущимся объектом и фиксированным центром проекции:

- найти пересечение линии взгляда и плоскости проекции;
- перенести точку пересечения в начало координат;
- повернуть вектор взгляда таким образом, чтобы он совместился с осью $+z$ и был направлен к началу системы координат (см. разд. 3-9);
- применить к сцене полученное преобразование;
- выполнить одноточечное перспективное проецирование на плоскость $z = 0$ из преобразованного центра проекции, расположенного теперь на оси z .

В качестве иллюстрации этого метода приведем относительно простой пример.

Пример 3-29 Перспективное проецирование на перпендикулярную плоскость проекции для фиксированного объекта

Рассмотрим обсуждавшийся ранее в примере 3-10 куб с отсеченным углом. Спроектируем куб из центра в точке $[10 10 10]$ на плоскость, проходящую через точку $[-1 -1 -1]$ и перпендикулярную направлению взгляда, как это показано на рис. 3-47.

Уравнение плоскости проекции можно получить из ее нормали (другие методы можно найти в [3-1]). В данном случае нормаль направлена в сторону, противоположную вектору взгляда.

Направление вектора взгляда

$$[s] = [-1 \quad -1 \quad -1].$$

Тогда нормаль к плоскости проекции, проходящей через точку $[-1 -1 -1]$ перпендикулярно направлению взгляда, будет

$$[n] = [1 \quad 1 \quad 1].$$

Общее выражение уравнения плоскости имеет вид:

$$ax + by + cz + d = 0.$$

Нормаль к такой плоскости задается вектором

$$[n] = [a \quad b \quad c].$$

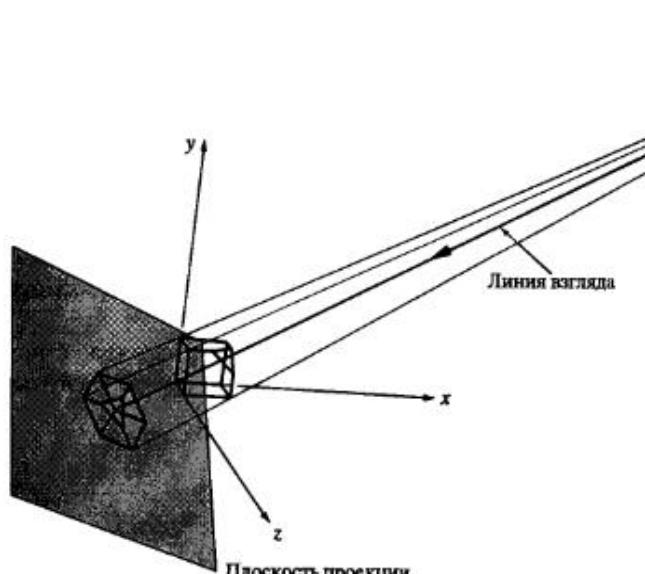


Рис. 3-47 Перспективное проецирование с перемещающимся центром проекции.

Значение d в уравнении плоскости вычисляется по любой точке, принадлежащей этой плоскости. Тогда уравнение плоскости проекции, проходящей через точку $[-1 - 1 - 1]$, примет вид:

$$x + y + z + d = 0$$

и

$$d = -x - y - z = 1 + 1 + 1 = 3.$$

Следовательно,

$$x + y + z + 3 = 0$$

есть уравнение плоскости проекции.

Пересечение направления взгляда и плоскости проекции определяется с помощью записи уравнения Прямой в параметрическом виде, подстановки его в уравнение плоскости и решения полученного уравнения.

Параметрическое уравнение вектора взгляда таково:

$$\begin{aligned} [S(t)] &= [x(t) \quad y(t) \quad z(t)] = \\ &= [10 \quad 10 \quad 10] + [-11 \quad -11 \quad -11]t, \quad 0 \leq t \leq 1. \end{aligned}$$

Его подстановка в уравнение плоскости приводит к

$$x(t) + y(t) + z(t) + 3 = (10 - 11t) + (10 - 11t) + (10 - 11t) + 3 = 0.$$

Решение уравнения дает значение параметра для точки пересечения, т. е.

$$-33t + 33 = 0 \rightarrow t = 1.0.$$

Саму точку пересечения получаем подстановкой t в $[S(t)]$. А именно,

$$\begin{aligned} [I] &= [S(1)] = [10 \quad 10 \quad 10] + [-11 \quad -11 \quad -11](1.0) = \\ &= [-1 \quad -1 \quad -1]. \end{aligned}$$

Как и ожидалось из простых геометрических рассуждений, точка пересечения находится в $x = y = z = -1$.

Требуемая матрица переноса (для переноса точки пересечения в начало координат) следующая:

$$[Tr] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

После переноса центр проекции находится в точке $[11 \quad 11 \quad 11]$, а направление взгляда проходит через начало координат.

Используя результаты разд. 3-9, совместить вектор взгляда с осью z можно с помощью поворота вокруг оси x на угол $\alpha = 45^\circ$ и последующего поворота вокруг оси y на угол $\beta = 35.26^\circ$. Матрицы поворотов записываются в виде:

$$[R_x] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{и} \quad [R_y] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ 0 & 1 & 0 & 0 \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

а суммарная матрица преобразования имеет вид

$$[M] = [Tr][R_x][R_y] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} & 0 \\ 0 & 0 & 3/\sqrt{3} & 1 \end{bmatrix}.$$

Преобразование центра проекции дает

$$\begin{aligned} [C_p][M] &= [10 \quad 10 \quad 10 \quad 1] \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} & 0 \\ 0 & 0 & 3/\sqrt{3} & 1 \end{bmatrix} = \\ &= [0 \quad 0 \quad 33/\sqrt{3} \quad 1]. \end{aligned}$$

Преобразование одноточечного центрального проецирования с центром в точке $z = 33/\sqrt{3}$ на плоскость $z = 0$ задается матрицей

$$[P_{rz}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\sqrt{3}/33 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Объединив с $[M]$, получим, что

$$\begin{aligned}[T] &= [M][P_{rz}] = \begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} & 0 \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} & 0 \\ 0 & 0 & 3/\sqrt{3} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\sqrt{3}/33 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 2/\sqrt{6} & 0 & 0 & -1/33 \\ -1/\sqrt{6} & 1/\sqrt{2} & 0 & -1/33 \\ -1/\sqrt{6} & -1/\sqrt{2} & 0 & -1/33 \\ 0 & 0 & 0 & 30/33 \end{bmatrix}.\end{aligned}$$

Преобразованные обычные координаты спроектированного объекта будут

$$\begin{aligned}[X^*] &= [X][T] = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0.5 & 1 & 1 \\ 0.5 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} 0.816 & 0 & 0 & -0.030 \\ -0.408 & 0.707 & 0 & -0.030 \\ -0.408 & -0.707 & 0 & -0.030 \\ 0 & 0 & 0 & 0.909 \end{bmatrix} = \\ &= \begin{bmatrix} -0.408 & -0.707 & 0 & 0.879 \\ 0.408 & -0.707 & 0 & 0.848 \\ 0.204 & -0.354 & 0 & 0.833 \\ -0.408 & 0 & 0 & 0.833 \\ -0.816 & 0 & 0 & 0.848 \\ 0 & 0 & 0 & 0.909 \\ 0.816 & 0 & 0 & 0.879 \\ 0.408 & 0.707 & 0 & 0.848 \\ -0.408 & 0.707 & 0 & 0.879 \\ 0.204 & 0.354 & 0 & 0.833 \end{bmatrix} = \\ &= \begin{bmatrix} -0.465 & -0.805 & 0 & 1 \\ 0.481 & -0.833 & 0 & 1 \\ 0.245 & -0.424 & 0 & 1 \\ -0.490 & 0 & 0 & 1 \\ -0.962 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0.929 & 0 & 0 & 1 \\ 0.481 & 0.833 & 0 & 1 \\ -0.465 & 0.805 & 0 & 1 \\ 0.245 & 0.424 & 0 & 1 \end{bmatrix}.\end{aligned}$$

Результат изображен на рис. 3-48.

3-21 ВОССТАНОВЛЕНИЕ ТРЕХМЕРНЫХ ОБЪЕКТОВ ПО ПРОЕКЦИЯМ

Задача восстановления трехмерной формы объекта или координат в пространстве широко распространена. Например, она постоянно возникает при использовании

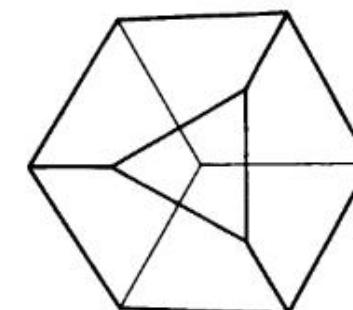


Рис. 3-48 Результат проецирования в примере 3-29.

чертежей, являющихся ортографическими проекциями. Хорошо известен метод восстановления трехмерного объекта или координаты по двум или более представленным на чертеже видам (ортографическим проекциям). Метод же восстановления трехмерного координатного вектора по двум перспективным проекциям, например по двум фотографиям, не столь хорошо известен. Конечно, если метод пригоден для перспективных проекций, то он подходит также и для более простых ортографических проекций и фактически для всех рассмотренных в предыдущих разделах проекций. Как будет показано ниже, при наличии определенной дополнительной информации не надо иметь точную информацию о самом преобразовании.

Прежде чем рассматривать задачу в общем виде, обратимся к специальному случаю восстановления трехмерных координат точки по двум или более ортографическим проекциям. На рис. 3-49 представлены ортографические виды (проекции) спереди, справа и сверху. При определении трехмерных координат точки A вид спереди дает значения координат x и y , вид справа — y и z и вид сверху — x и z , т. е.

$$\begin{array}{ll} \text{спереди (front): } & x_f \quad y_f \\ \text{справа (right side): } & y_r \quad z_r \\ \text{сверху (top): } & x_t \quad z_t \end{array}$$

Заметим, что для каждой координаты мы получаем два значения. В математике в подобном случае говорят, что на задачу наложены избыточные условия. Ведь надо найти только три независимые величины, тогда как известны шесть определяющих их условий (уравнений). Для любой измерительной системы в общем случае $x_f \neq x_t$, $y_f \neq y_r$, $z_r \neq z_t$ ¹. Так как неизвестно, какое из значений верно, то наиболее разумным решением будет усреднить их.

Вернувшись теперь к восстановлению трехмерных координат по перспективным проекциям, напомним, что обобщенное перспективное преобразование представляется в виде матриц размера 4×4 . Таким образом,

$$[x \ y \ z \ 1][T'] = [x' \ y' \ z' \ h],$$

¹ По этой причине в чертежах размеры указываются явным образом.

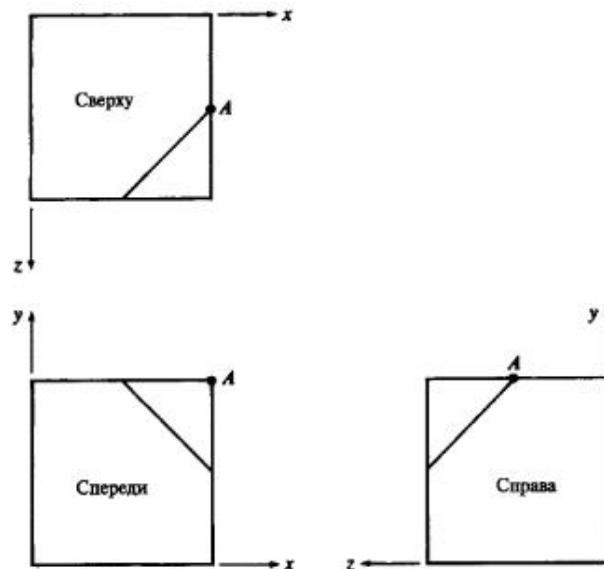


Рис. 3-49 Восстановление трехмерной формы по ортографическим проекциям.

где

$$[T'] = \begin{bmatrix} T'_{11} & T'_{12} & T'_{13} & T'_{14} \\ T'_{21} & T'_{22} & T'_{23} & T'_{24} \\ T'_{31} & T'_{32} & T'_{33} & T'_{34} \\ T'_{41} & T'_{42} & T'_{43} & T'_{44} \end{bmatrix}.$$

Результаты могут быть спроектированы на двумерную плоскость, скажем на плоскость $z = 0$, с помощью

$$[T''] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Объединение двух матриц дает

$$[T] = [T''] [T'] = \begin{bmatrix} T_{11} & T_{12} & 0 & T_{14} \\ T_{21} & T_{22} & 0 & T_{24} \\ T_{31} & T_{32} & 0 & T_{34} \\ T_{41} & T_{42} & 0 & T_{44} \end{bmatrix}.$$

Полезно записать это преобразование в виде

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} T_{11} & T_{12} & 0 & T_{14} \\ T_{21} & T_{22} & 0 & T_{24} \\ T_{31} & T_{32} & 0 & T_{34} \\ T_{41} & T_{42} & 0 & T_{44} \end{bmatrix} = [x' & y' & 0 & h] = h[x^* & y^* & 0 & 1]. \quad (3-71)$$

Заметим, что x^* и y^* — это координаты перспективной проекции на плоскость $z = 0$. Могли бы также использоваться проекции на плоскости $x = 0$ или $y = 0$.

Расписав в явном виде матричное уравнение (3-71), получим

$$T_{11}x + T_{21}y + T_{31}z + T_{41} = hx^*, \quad (3-72\text{a})$$

$$T_{12}x + T_{22}y + T_{32}z + T_{42} = hy^*, \quad (3-72\text{b})$$

$$T_{14}x + T_{24}y + T_{34}z + T_{44} = h. \quad (3-72\text{c})$$

Подставим значение h из (3-72 c) в уравнения (3-72 a) и (3-72 b):

$$(T_{11} - T_{14}x^*)x + (T_{21} - T_{24}x^*)y + (T_{31} - T_{34}x^*)z + (T_{41} - T_{44}x^*) = 0, \quad (3-73\text{a})$$

$$(T_{12} - T_{14}y^*)x + (T_{22} - T_{24}y^*)y + (T_{32} - T_{34}y^*)z + (T_{42} - T_{44}y^*) = 0. \quad (3-73\text{b})$$

По предложению Сазерленда [3-4] эту пару уравнений можно рассматривать тремя разными способами. В первом предполагаются известными T , x , y , z . Тогда у нас есть два уравнения с двумя неизвестными x^* и y^* . Таким образом, эти уравнения можно использовать для непосредственного вычисления координат перспективной проекции. Именно этот подход применялся во всех предыдущих рассуждениях в данной главе.

При другом подходе полагаются известными T , x^* , y^* . В этом случае получается два уравнения от трех неизвестных пространственных координат x , y , z . Этую систему уравнений невозможно решить. Однако, если известны две перспективные проекции, скажем две фотографии, тогда уравнения (3-73) можно записать для обеих проекций. В результате получим

$$(T_{11}^1 - T_{14}^1x^{*1})x + (T_{21}^1 - T_{24}^1x^{*1})y + (T_{31}^1 - T_{34}^1x^{*1})z + (T_{41}^1 - T_{44}^1x^{*1}) = 0,$$

$$(T_{12}^1 - T_{14}^1y^{*1})x + (T_{22}^1 - T_{24}^1y^{*1})y + (T_{32}^1 - T_{34}^1y^{*1})z + (T_{42}^1 - T_{44}^1y^{*1}) = 0,$$

$$(T_{11}^2 - T_{14}^2x^{*2})x + (T_{21}^2 - T_{24}^2x^{*2})y + (T_{31}^2 - T_{34}^2x^{*2})z + (T_{41}^2 - T_{44}^2x^{*2}) = 0,$$

$$(T_{12}^2 - T_{14}^2y^{*2})x + (T_{22}^2 - T_{24}^2y^{*2})y + (T_{32}^2 - T_{34}^2y^{*2})z + (T_{42}^2 - T_{44}^2y^{*2}) = 0,$$

где верхние индексы 1 и 2 обозначают первую и вторую перспективные проекции. Заметим, что преобразования $[T^1]$ и $[T^2]$ не обязательно должны быть одинаковыми. Эти уравнения можно переписать в матричной форме

$$[A][X] = [B], \quad (3-74)$$

где

$$A = \begin{bmatrix} T_{11}^1 - T_{14}^1x^{*1} & T_{21}^1 - T_{24}^1x^{*1} & T_{31}^1 - T_{34}^1x^{*1} \\ T_{12}^1 - T_{14}^1y^{*1} & T_{22}^1 - T_{24}^1y^{*1} & T_{32}^1 - T_{34}^1y^{*1} \\ T_{11}^2 - T_{14}^2x^{*2} & T_{21}^2 - T_{24}^2x^{*2} & T_{31}^2 - T_{34}^2x^{*2} \\ T_{12}^2 - T_{14}^2y^{*2} & T_{22}^2 - T_{24}^2y^{*2} & T_{32}^2 - T_{34}^2y^{*2} \end{bmatrix},$$

$$[X]^T = [x \ y \ z]$$

$$[B]^T = [T_{44}^1x^{*1} - T_{41}^1 \ T_{44}^1y^{*1} - T_{42}^1 \ T_{44}^2x^{*2} - T_{41}^2 \ T_{44}^2y^{*2} - T_{42}^2].$$

Уравнения (3-73) представляют четыре уравнения от трех неизвестных пространственных координат x, y, z . Для нахождения решения $[X]$ матрица $[A]$ не может быть инвертирована, так как она не квадратная. Снова, как и в случае восстановления трехмерных координат по ортографическим проекциям, условия избыточны и, таким образом, задача может быть решена только в некотором усредненном или наиболее подходящем смысле.

Усредненное решение можно вычислить, вспомнив, что матрица, умноженная на свою транспонированную матрицу, является квадратной. Таким образом, умножив обе части уравнения (3-74) на $[A]^T$, получим

$$[A]^T [A] [X] = [A]^T [B].$$

Вычислив обратную матрицу $[[A]^T [A]]^{-1}$, получим усредненное решение для $[X]$, т.е.

$$[X] = [[A]^T [A]]^{-1} [A]^T [B]. \quad (3-75)$$

Если решения для $[X]$ нет, то наложенные условия избыточны и не существует единственного решения, обеспечивающего наименьшую ошибку. Этот метод иллюстрируется на примере.

Пример 3-30 Трехмерное восстановление

Предположим, что обмеренные координаты точки на одной перспективной проекции равны $[0.836 \ 1.836 \ 0 \ 1]$ и $[0.6548 \ 0 \ 0.2886 \ 1]$ на другой. Первое преобразование перспективной проекции является результатом поворота на 60° вокруг оси y и последующего переноса на 2 единицы в отрицательном направлении оси y . Точка проекции находится в $z = -1$, а результат проецируется на плоскость $z = 0$. Это действительно является двуточечной перспективной проекцией. Второе преобразование является результатом поворотов на 30° вокруг каждой из осей x, y . Точка проекции находится в $y = -1$, а результат проецируется на плоскость $y = 0$, т.е. является трехточечной перспективной проекцией. Таким образом преобразования $[T^1]$ и $[T^2]$

$$[T^1] = \begin{bmatrix} 0.5 & 0 & 0 & -0.87 \\ 0 & 1 & 0 & 0 \\ 0.87 & 0 & 0 & 0.5 \\ 0 & -2 & 0 & 1 \end{bmatrix} \text{ и } [T^2] = \begin{bmatrix} 0.87 & 0 & -0.5 & 0 \\ 0.25 & 0 & 0.43 & 0.87 \\ 0.43 & 0 & 0.75 & -0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Отметим, что последние две строки $[A]$ и $[B]$ должны быть переписаны, чтобы учсть тот факт, что $[T^2]$ является проецированием на плоскость $y = 0$. Тогда матрица A

$$[A] = \begin{bmatrix} 1.22 & 0 & 0.45 \\ -1.59 & 1 & 0.92 \\ 0.87 & -0.32 & 0.76 \\ -0.5 & 0.18 & 0.89 \end{bmatrix}$$

и

$$[B]^T = [0.84 \ 0.16 \ 0.65 \ 0.29].$$

Решив уравнение, получим $[X] = [0.5 \ 0.5 \ 0.5]$, т.е. центр единичного куба.

В качестве третьего подхода к рассмотрению уравнений (3-73) отметим, что если известны координаты нескольких точек в объектном пространстве и на перспективной проекции, то можно найти элементы преобразования T_{ij} . Эти элементы впоследствии можно использовать для нахождения местоположения неизвестных точек, используя описанный выше второй метод. Чтобы показать это, перепишем (3-73) в виде

$$T_{11}x + T_{21}y + T_{31}z + T_{41} - T_{14}xx^* - T_{24}yx^* - T_{34}zx^* - T_{44}x^* = 0, \quad (3-76 \text{ a})$$

$$T_{12}x + T_{22}y + T_{32}z + T_{42} - T_{14}xy^* - T_{24}yy^* - T_{34}zy^* - T_{44}y^* = 0. \quad (3-76 \text{ b})$$

В предположении, что известны x^*, y^*, x, y, z , уравнения (3-76 a) и (3-76 b) представляют два уравнения с 12 неизвестными элементами T_{ij} . Применяя эти уравнения к 6 некомпланарным точкам в объектном пространстве и на перспективной проекции, получим систему из 12 уравнений с 12 неизвестными. Эти уравнения могут быть решены относительно T_{ij} . Таким образом, находится преобразование, породившее перспективную проекцию, например фотографию. Заметим, что в этом случае не требовалось никакой предварительной информации о преобразовании. Если, например, это были фотографии, то не требовалось знать ни о местоположении, ни об ориентации фотокамеры. В матричном виде система уравнений записывается в таком виде:

$$\begin{bmatrix} x_1 & 0 & -x_1x_1^* & y_1 & 0 & -y_1x_1^* & z_1 & 0 & -z_1x_1^* & 1 & 0 & -x_1^* \\ 0 & x_1 & -x_1y_1^* & 0 & y_1 & -y_1y_1^* & 0 & z_1 & -z_1y_1^* & 0 & 1 & -y_1^* \\ x_2 & 0 & -x_2x_2^* & y_2 & 0 & -y_2x_2^* & z_2 & 0 & -z_2x_2^* & 1 & 0 & -x_2^* \\ 0 & x_2 & -x_2y_2^* & 0 & y_2 & -y_2y_2^* & 0 & z_2 & -z_2y_2^* & 0 & 1 & -y_2^* \\ x_3 & 0 & -x_3x_3^* & y_3 & 0 & -y_3x_3^* & z_3 & 0 & -z_3x_3^* & 1 & 0 & -x_3^* \\ 0 & x_3 & -x_3y_3^* & 0 & y_3 & -y_3y_3^* & 0 & z_3 & -z_3y_3^* & 0 & 1 & -y_3^* \\ x_4 & 0 & -x_4x_4^* & y_4 & 0 & -y_4x_4^* & z_4 & 0 & -z_4x_4^* & 1 & 0 & -x_4^* \\ 0 & x_4 & -x_4y_4^* & 0 & y_4 & -y_4y_4^* & 0 & z_4 & -z_4y_4^* & 0 & 1 & -y_4^* \\ x_5 & 0 & -x_5x_5^* & y_5 & 0 & -y_5x_5^* & z_5 & 0 & -z_5x_5^* & 1 & 0 & -x_5^* \\ 0 & x_5 & -x_5y_5^* & 0 & y_5 & -y_5y_5^* & 0 & z_5 & -z_5y_5^* & 0 & 1 & -y_5^* \\ x_6 & 0 & -x_6x_6^* & y_6 & 0 & -y_6x_6^* & z_6 & 0 & -z_6x_6^* & 1 & 0 & -x_6^* \\ 0 & x_6 & -x_6y_6^* & 0 & y_6 & -y_6y_6^* & 0 & z_6 & -z_6y_6^* & 0 & 1 & -y_6^* \end{bmatrix} = 0, \quad (3-77)$$

где нижние индексы соответствуют точкам с известным местоположением. Уравнения (3-77) записываются в более компактном виде:

$$[A'][T] = 0.$$

Так как уравнения (3-77) являются однородными уравнениями, то они содержат произвольный масштабный коэффициент. Следовательно, T_{44} можно приравнять единице и нормализовать результирующее преобразование. Система сводится к 11 уравнениям или 5 1/2 точки. Если преобразование нормализовано, то

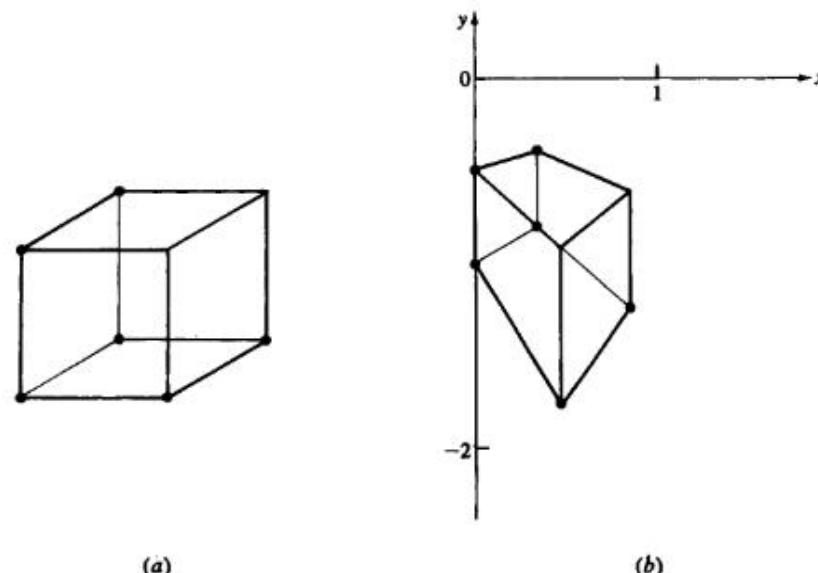


Рис. 3-50 Восстановление преобразования по перспективной проекции.

последний столбец в $[A']$ переносится в правую часть и решается неоднородное матричное уравнение. Ниже приводится пример.

Пример 3-31 Основы восстановления

В качестве конкретного примера рассмотрим единичный куб с шестью известными угловыми точками.

$$[P] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Как показано на рис. 3-50, соответствующие им точки на проекции отмечены жирными кружками. Преобразованные координаты этих точек таковы:

$$\begin{bmatrix} 0 & -1 \\ 0.34 & -0.8 \\ 0.34 & -0.4 \\ 0 & -0.5 \\ 0.44 & -1.75 \\ 0.83 & -1.22 \end{bmatrix}$$

Тогда (3-77) записывается в виде

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.34 & 1 & 0 & -0.34 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.8 & 0 & 1 & 0.8 \\ 0 & 0 & 0 & 1 & 0 & -0.34 & 1 & 0 & -0.34 & 1 & 0 & -0.34 \\ 0 & 0 & 0 & 0 & 1 & 0.4 & 0 & 1 & 0.4 & 0 & 1 & 0.4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.5 & 0 & 0 & 0 & 0 & 1 & 0.5 \\ 1 & 0 & -0.44 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.44 \\ 0 & 1 & 1.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1.75 \\ 1 & 0 & -0.83 & 0 & 0 & 0 & 1 & 0 & -0.83 & 1 & 0 & -0.83 \\ 0 & 1 & 1.22 & 0 & 0 & 0 & 0 & 1 & 1.22 & 0 & 1 & 1.22 \end{bmatrix} \begin{bmatrix} T_{11} \\ T_{12} \\ T_{14} \\ T_{21} \\ T_{22} \\ T_{24} \\ T_{31} \\ T_{32} \\ T_{34} \\ T_{41} \\ T_{42} \\ T_{44} \end{bmatrix} = 0.$$

Решение относительно 12 неизвестных элементов T_{ij} такое:

$$\begin{bmatrix} T_{11} \\ T_{12} \\ T_{14} \\ T_{21} \\ T_{22} \\ T_{24} \\ T_{31} \\ T_{32} \\ T_{34} \\ T_{41} \\ T_{42} \\ T_{44} \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0 \\ -0.43 \\ 0 \\ 0.5 \\ 0 \\ 0.43 \\ 0 \\ 0.25 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Подстановка этих результатов в матрицу $[T]$ 4×4 дает

$$[T] = \begin{bmatrix} 0.25 & 0 & 0 & -0.43 \\ 0 & 0.5 & 0 & 0 \\ 0.43 & 0 & 0 & 0.25 \\ 0 & -1 & 0 & 1 \end{bmatrix}.$$

3-22 ЛИТЕРАТУРА

- 3-1 Rogers, D. F., *Procedural Elements for Computer Graphics*, McGraw-Hill Book Co., New York, 1985. (См. [1-1] на с. 74.)
- 3-2 Carl bom, I., and Paciorek, J., «Planar Geometric Projections and Viewing Transformations», *ACM Comp. Surv.*, Vol. 10, No. 4, pp. 465–502, 1978.
- 3-3 Sutherland, I. E., «Three Dimensional Data Input by Tablet», *Proc. IEEE*, Vol. 62, No. 2, pp. 453–461, 1974.
- 3-4 Slama, C. (ed.), *Manual of Photogrammetry*, American Society of Photogrammetry, 1980.
- 3-5 Kingslake, R. (ed.), *Applied Optics and Optical Engineering*, Vol. 2, Academic Press, New York, 1965.

ПЛОСКИЕ КРИВЫЕ

4-1 ВВЕДЕНИЕ

Существует множество способов построения кривых вручную с помощью карандаша, ручки, кисточки, ножа и разнообразных инструментов: линейки, лекала, циркуля, плазы, шаблона и т. д. Каждый инструмент служит определенной цели, причем нет ни одного абсолютного универсального. Точно так же в машинной графике кривые строятся с помощью разных методов и инструментов. В этой главе рассматриваются методы построения двумерных кривых, т. е. полностью лежащих в одной плоскости. Здесь мы ограничимся коническими сечениями.

4-2 ПРЕДСТАВЛЕНИЕ КРИВЫХ

В предыдущих двух главах изучались преобразования точек. Кривая может быть представлена совокупностью точек. Если точки расположены близко друг от друга, то, соединяя их отрезками прямой, мы получим изображение кривой. На рис. 4.1 показаны два разных точечных представления одной кривой. Точки на кривой 4.1a расположены равномерно по ее длине. В результате получается довольно плохое представление кривой, особенно если мал радиус кривизны. Улучшить вид можно, увеличивая плотность точек в таких участках, рис. 4.1b.

Хотя, как показано выше, кривые могут быть достаточно хорошо представлены совокупностью точек, аналитическое представление во многих отношениях предпочтительнее. Его преимущества — это точность, компактность записи и простота вычисления промежуточных точек. Аналитическое представление позволяет без труда определить наклон и радиус кривизны, а при точечном представле-

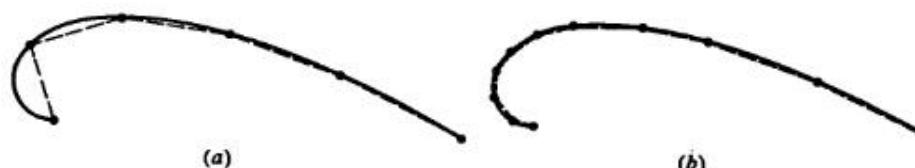


Рис. 4.1 Точечное представление кривых. (a) Равномерная плотность точек вдоль кривой; (b) плотность точек возрастает с уменьшением радиуса кривизны.

ни для этого требуется численное дифференцирование — чрезвычайно неточная процедура. Сравните объем памяти для точечного представления окружности с 32 точками на ней и аналитическое представление — координаты центра и радиус. При аналитическом представлении кривой можно точно определить положение любой точки, а при точечном нужна интерполяция, причем в общем случае результат интерполяции не принадлежит кривой.

Опыт показывает, что аналитически представленные кривые легче изобразить на рисунке. Аналитическое представление выгоднее, когда для удовлетворения заданному критерию необходимо непрерывно изменять форму кривой. Ниже рассматриваются различные методы аналитического представления двумерных кривых.

Часто требуется аналитическое представление кривой, первоначально заданной точками. С математической точки зрения это проблема интерполяции. Для того чтобы провести кривую через все заданные точки, применяется метод кусочной полиномиальной аппроксимации. Для этого требуется определить коэффициенты полинома некоторой степени. Вид кривой в промежуточных точках зависит от степени полинома и граничных условий. Метод рассматривается в гл. 5.

Если же точки — только приблизительные значения величин, например данные экспериментальных измерений или наблюдений, то нужна кривая, задающая верное направление изменения. В общем случае кривая может не проходить ни через одну точку данных. Здесь применяется метод наименьших квадратов [4-1]. Метод наименьших квадратов дает кривую в виде $y = f(x)$, которая минимизирует сумму квадратов отклонений по оси y между данными и полученной кривой. Вид $y = f(x)$ выбирают, исходя из характера изучаемого процесса.

Обычно для метода наименьших квадратов используются полиномы, степенные и экспоненциальные функции: $y = ax^b$, $y = ae^{bx}$ или $y = c_1 + c_2x + c_3x^2 + \dots + c_{n+1}x^n$, где a , b и c_i — константы. Независимо от выбора вида кривой метод требует решения системы линейных алгебраических уравнений для определения неизвестных констант. Более подробное описание приведено в работе [4-1] и книгах по численному анализу.

4-3 НЕПАРАМЕТРИЧЕСКИЕ КРИВЫЕ

Математически кривая может быть представлена в параметрической или непараметрической форме. Непараметрическая кривая задается в виде явной или неявной функции. Для плоской кривой явное непараметрическое представление имеет вид:

$$y = f(x).$$

Пример — уравнение прямой, $y = mx + b$. При этом одному значению x соответствует только одно значение y , поэтому замкнутые или многозначные кривые, например окружность, явно представить нельзя. Неявное представление

$$f(x, y) = 0$$

позволяет обойти это ограничение.

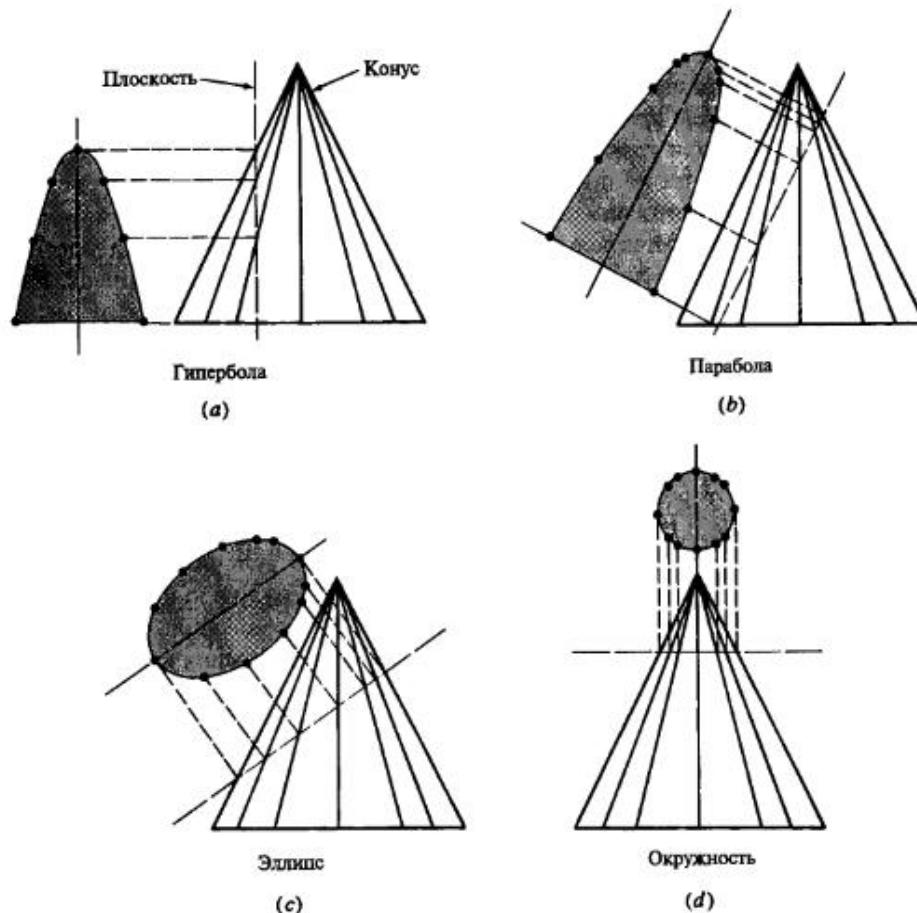


Рис. 4-2 Конические сечения.

Общий вид неявного уравнения второй степени

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0$$

порождает различные двумерные кривые, называемые коническими сечениями. На рис. 4.2 изображены три вида конических сечений — парабола, гипербола и эллипс. Окружность — это частный случай эллипса. Определяя коэффициенты a , b , c , d , e и f , можно получить разные конические сечения. Если сечение задано относительно локальной системы координат и проходит через ее начало, то $f = 0$. Для того чтобы провести кривую через данные точки, используются граничные условия.

Пусть $c = 1.0$, тогда сегмент кривой между двумя точками определяется пятью независимыми условиями, из которых вычисляются оставшиеся коэффициенты a , b , d , e и f . Например, можно указать положение крайних точек, наклон кривой в них и промежуточную точку на кривой.

4-4 Параметрические кривые

Если $b = 0$ и $c = 1.0$, то аналитическое представление кривой получается с помощью только четырех дополнительных условий, например положения концевых точек и наклона кривой в них. Кривая при $a = 1.0$, $b = 0$ и $c = 1.0$ еще проще:

$$x^2 + y^2 + 2dx + 2ey + f = 0.$$

Тремя условиями для вычисления d , e и f могут быть две концевые точки и наклон кривой в одной из них или же две концевые точки и третья точка на кривой.

При $a = b = c = 0$ получается прямая линия. Ее уравнение

$$dx + ey = f = 0$$

или

$$y = -\left(\frac{d}{e}\right)x - \frac{f}{e} = mx + b',$$

где m — наклон линии, b' — пересечение с осью y .

Как явное, так и неявное непараметрическое представление осезависимо, т. е. сложность обработки зависит от выбора системы координат. Например, если в заданной системе координат граничным условием является вертикальный наклон, нужно либо изменить ее, либо аппроксимировать бесконечный коэффициент наклона наибольшей допустимой положительной или отрицательной величиной.

Кроме того, если точки на осезависимой непараметрической кривой вычисляются с равномерным приращением по x или y , они не будут равномерно распределены вдоль кривой. Это может повлиять на качество и точность графического изображения. Тем не менее непараметрическое представление бывает полезным. Теперь рассмотрим параметрическое представление, позволяющее обойти эти ограничения.

4-4 ПАРАМЕТРИЧЕСКИЕ КРИВЫЕ

В параметрическом виде каждая координата точки кривой представлена как функция одного параметра. Значение параметра задает координатный вектор точки на кривой. Для двумерной кривой с параметром t координаты точки равны:

$$x = x(t),$$

$$y = y(t).$$

Тогда векторное представление точки на кривой:

$$P(t) = [x(t) \quad y(t)].$$

Чтобы получить непараметрическую форму, нужно исключить t из двух уравнений и вывести одно в терминах x и y .

Параметрическая форма позволяет представить замкнутые и многозначные кривые. Производная, т. е. касательный вектор, есть

$$P'(t) = [x'(t) \quad y'(t)],$$

где ' обозначает дифференцирование по параметру. Наклон кривой, dy/dx , равен

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)}.$$

Отметим, что при $x'(t) = 0$ наклон бесконечен. Параметрическое представление не вызывает в этом случае вычислительных трудностей, достаточно приравнять нулю одну компоненту касательного вектора.

Так как точка на параметрической кривой определяется только значением параметра, эта форма не зависит от выбора системы координат. Конечные точки и длина кривой определяются диапазоном изменения параметра. Часто бывает удобно нормализовать параметр на интересующем отрезке кривой к $0 \leq t \leq 1$. Осенезависимость параметрической кривой позволяет с легкостью проводить с ней аффинные преобразования, рассмотренные в гл. 2 и 3.

Самое простое параметрическое представление у прямой. Для двух векторов положения P_1 и P_2 параметрический вид отрезка прямой между ними такой:

$$P(t) = P_1 + (P_2 - P_1)t, \quad 0 \leq t \leq 1.$$

Так как $P(t)$ это вектор, у каждой его составляющей есть параметрическое представление $x(t)$ и $y(t)$ между P_1 и P_2 :

$$x(t) = x_1 + (x_2 - x_1)t, \quad 0 \leq t \leq 1.$$

$$y(t) = y_1 + (y_2 - y_1)t,$$

Пример 4-1 Параметрическое представление прямой

Найти параметрическое представление отрезка между точками $P_1[1 \ 2]$ и $P_2[4 \ 3]$, касательный вектор и наклон.

Параметрическое представление:

$$P(t) = P_1 + (P_2 - P_1)t = [1 \ 2] + ([4 \ 3] - [1 \ 2])t, \quad 0 \leq t \leq 1,$$

$$P(t) = [1 \ 2] + [3 \ 1]t, \quad 0 \leq t \leq 1.$$

Параметрическое представление составляющих x и y :

$$x(t) = x_1 + (x_2 - x_1)t = 1 + 3t, \quad 0 \leq t \leq 1,$$

$$y(t) = y_1 + (y_2 - y_1)t = 2 + t.$$

Дифференцируя $P(t)$, получим касательный вектор:

$$P'(t) = [x'(t) \ y'(t)] = [3 \ 1]$$

или

$$\bar{V}_t = 3\mathbf{i} + \mathbf{j},$$

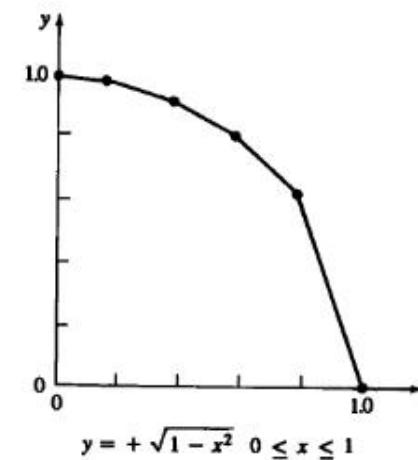
где \bar{V}_t — касательный вектор, а \mathbf{i}, \mathbf{j} — единичные векторы в направлениях x , y соответственно.

Наклон отрезка равен

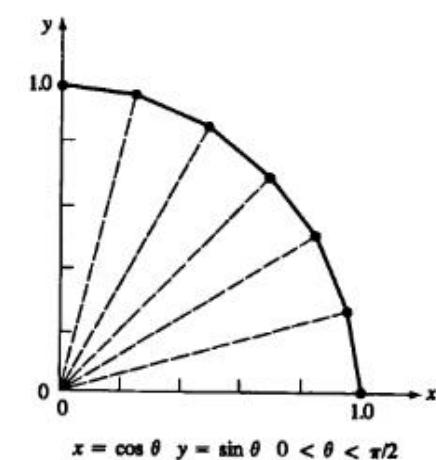
$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)} = \frac{1}{3}.$$

На рис. 4.3 сравниваются непараметрическое и параметрическое представления окружности в первом квадранте. Непараметрический вид

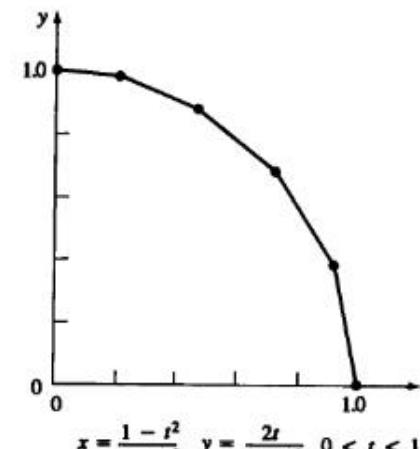
$$y = +\sqrt{1 - x^2}, \quad 0 \leq x \leq 1, \quad (4-1)$$



(a)



(b)



(c)

Рис. 4-3 Представление окружности для первого квадранта.

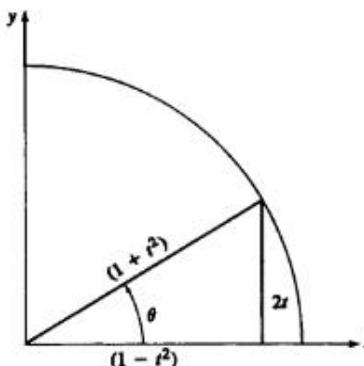


Рис. 4-4 Связь между параметрическими представлениями.

показан на рис. 4.3 a. Точки на дуге соответствуют равным приращениям x . При этом дуга состоит из отрезков разной длины, и получается весьма приблизительное графическое представление окружности. Кроме того, расчет квадратного корня — вычислительно дорогостоящая операция.

Стандартная параметрическая форма единичной окружности:

$$\begin{aligned} x &= \cos \theta, \quad 0 \leq \theta \leq 2\pi, \\ y &= \sin \theta, \end{aligned}$$

или

$$P(\theta) = [x \ y] = [\cos \theta \ \sin \theta], \quad 0 \leq \theta \leq 2\pi, \quad (4-2)$$

где параметр θ — геометрический угол, отмеряемый против часовой стрелки от положительной полусоси x . На рис. 3.4 b изображена дуга, построенная по равным приращениям параметра в пределах $0 \leq \theta \leq \pi/2$. При этом точки располагаются на одинаковом расстоянии вдоль окружности, и окружность выглядит гораздо лучше. Недостаток такого представления — сложность вычисления тригонометрических функций. (Более простой метод рассматривается ниже в разд. 4.5.)

Параметрическое представление кривой не единственное, например,

$$P(t) = \left[\frac{(1-t^2)}{(1+t^2)} \quad \frac{2t}{(1+t^2)} \right], \quad 0 \leq t \leq 1 \quad (4-3)$$

также представляет дугу единичной окружности в первом квадранте (рис. 4.3 c). Связь между параметрическим представлением (3.4) и стандартным параметрическим представлением (4.2) показана на рис. 4.4. Из него видно, что для единичной окружности

$$\begin{aligned} x &= \cos \theta = \frac{1-t^2}{1+t^2}, \quad 0 \leq \theta \leq \pi/2, \quad 0 \leq t \leq 1, \\ y &= \sin \theta = \frac{2t}{1+t^2}, \quad 0 \leq \theta \leq \pi/2, \quad 0 \leq t \leq 1. \end{aligned}$$

Факт, что уравнение 4.3 представляет дугу единичной окружности, подтверждается следующим:

$$r^2 = x^2 + y^2 = \left(\frac{1-t^2}{1+t^2} \right)^2 + \left(\frac{2t}{1+t^2} \right)^2 = \frac{1-2t^2+t^4+4t^2}{(1+t^2)^2} = \frac{(1+t^2)^2}{(1+t^2)^2} = 1,$$

где r — единичный радиус.

На рис. 4-3 с показан результат для равных приращений t . Он лучше, чем у явного (4-1), но хуже, чем у стандартного параметрического представления (4-2). Однако уравнение (4-3) проще с вычислительной точки зрения, т. е. это компромиссное решение.

Пример 4-2 Определение точки на параметрической кривой

Пусть надо найти y для заданного значения x . Например, пусть $x = 0.5$, и требуется вычислить y для единичной окружности. Для явного представления (4-1) вычисляем непосредственно:

$$y = \sqrt{1-x^2} = \sqrt{1-(0.5)^2} = \sqrt{0.75} = 0.866.$$

Для параметрического представления сначала нужно выразить параметр t через x , а затем с помощью полученного значения найти y . Параметрическое представление уравнения (4-2)

$$\begin{aligned} x &= \cos \theta, \\ y &= \sin \theta. \end{aligned}$$

Поэтому

$$\begin{aligned} \theta &= \arccos(x) = \arccos(0.5) = 60^\circ, \\ y &= \sin(60^\circ) = 0.866. \end{aligned}$$

С другой стороны, из уравнения 4-3

$$\begin{aligned} x &= \frac{1-t^2}{1+t^2}, \\ y &= \frac{2t}{1+t^2}. \end{aligned}$$

Решаем первое из этих уравнений относительно

$$t = \left(\frac{1-x}{1+x} \right)^{1/2} = \frac{1}{\sqrt{3}} = 0.57735$$

и получаем

$$y = \frac{2/\sqrt{3}}{4/3} = \frac{\sqrt{3}}{2} = 0.866.$$

В случае более сложного параметрического представления бывает удобнее использовать значение явной переменной итеративными методами.

Параметрическое представление конических сечений осенезависимо и дает более качественное изображение, чем непараметрическое; однако оба имеют свои достоинства и недостатки и часто применяются в машинной графике.

4-5 ПАРАМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ОКРУЖНОСТИ

Параметрическое представление окружности радиуса r с центром в начале координат:

$$\begin{aligned}x &= r \cos \theta, & 0 \leq \theta \leq 2\pi, \\y &= r \sin \theta,\end{aligned}\quad (4-4)$$

где θ — параметр. Ранее мы отмечали, что равномерное приращение дает отличное изображение, но алгоритм неэффективен из-за частого вызова тригонометрических функций. В работе [4-2] приводится более быстрый алгоритм.

Полной окружности соответствует диапазон изменения параметра θ от 0 до 2π . Если рассматривать некоторое фиксированное число равномерно распределенных точек по окружности, то приращение параметра между точками можно считать константой. Координаты любой точки на окружности с центром в начале координат

$$\begin{aligned}x_{i+1} &= r \cos(\theta_i + \delta\theta), \\y_{i+1} &= r \sin(\theta_i + \delta\theta),\end{aligned}$$

где θ_i — значение параметра для точки x_i, y_i .

По формуле суммы углов

$$\begin{aligned}x_{i+1} &= r(\cos \theta_i \cos \delta\theta - \sin \theta_i \sin \delta\theta), \\y_{i+1} &= r(\cos \theta_i \sin \delta\theta + \sin \theta_i \cos \delta\theta).\end{aligned}$$

Вспомним уравнение (4-4) для $\theta = \theta_i$:

$$\begin{aligned}x_i &= r \cos \theta_i, \\y_i &= r \sin \theta_i\end{aligned}$$

и получим рекурсивные уравнения

$$\begin{aligned}x_{i+1} &= x_i \cos \delta\theta - y_i \sin \delta\theta, \\y_{i+1} &= x_i \sin \delta\theta + y_i \cos \delta\theta,\end{aligned}\quad (4-5)$$

что соответствует повороту точки x_i, y_i на $\delta\theta$.

Так как $\delta\theta$ постоянно и равно $2\pi/(n-1)$, где n — количество равномерно распределенных по окружности точек, значения \sin и \cos нужно вычислить только один раз. Во внутреннем цикле используются только четыре умножения, вычитание и сложение, поэтому алгоритм работает очень быстро¹. Результат на рис. 4-5 соответствует расчетам по формуле (4-4).

Окружность с центром в произвольной точке получается переносом окружности соответствующего радиуса с центром в начале координат. В некоторых случаях можно упростить задачу: сначала строить единичную окружность с центром в начале координат, а затем, комбинируя перенос и масштабирование, получить окружности с любым радиусом и центром.

¹ Замечание. Как и для всех рекурсивных алгоритмов результаты при больших n могут сильно зависеть от конкретных данных.

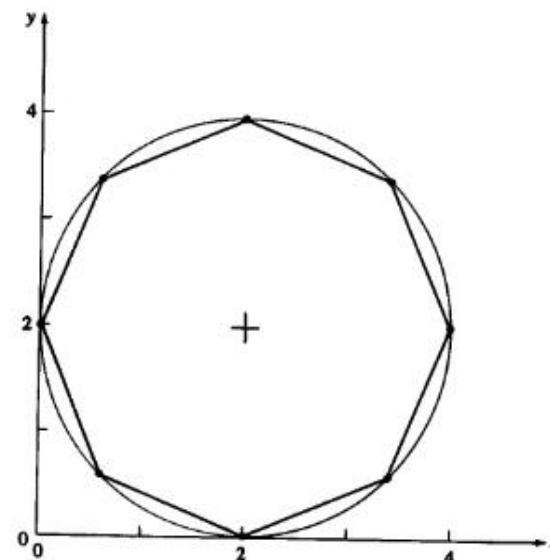


Рис. 4-5 Единичная параметрическая окружность с $n = 8$.

Пример 4-3 Генерация параметрической окружности

Построить окружность радиуса 2 с центром в точке (2, 2). Рассмотрим два метода. Первый: построить окружность радиуса 2 с центром в начале координат и перенести на 2 единицы по x и y . Второй: построить единичную окружность с центром в начале координат, увеличить в два раза и, наконец, перенести. Воспользуемся вторым методом. Пусть для простоты на окружности лежат восемь точек, хотя обычно требуется гораздо большее количества. На самом деле количество точек зависит от радиуса.

Окружность — это замкнутая кривая, поэтому первая ($\theta = 0$) и последняя ($\theta = 2\pi$) точки совпадают. Следовательно, чтобы получить n различных точек на окружности, нужно вычислить $n+1$ точку. Для незамкнутых кривых это необязательно.

Найдем $\delta\theta$:

$$\delta\theta = \frac{2\pi}{(n+1-1)} = \frac{2\pi}{n} = \frac{2\pi}{8} = \frac{\pi}{4}.$$

Исходные значения x и y из уравнения (4-4) при $\theta = 0$ таковы:

$$\begin{aligned}x_1 &= r \cos \theta_1 = (1) \cos(0) = 1, \\y_1 &= r \sin \theta_1 = (1) \sin(0) = 0.\end{aligned}$$

Теперь по формуле (4-5) получим остальные семь точек. Для первой

$$\begin{aligned}\sin \delta\theta &= \sin \frac{\pi}{4} = \frac{\sqrt{2}}{2}, \\\cos \delta\theta &= \cos \frac{\pi}{4} = \frac{\sqrt{2}}{2}\end{aligned}$$

Таблица 4-1 Результаты для единичной окружности

i	x_i	y_i
1	1	0
2	$\sqrt{2}/2$	$\sqrt{2}/2$
3	0	1
4	$-\sqrt{2}/2$	$\sqrt{2}/2$
5	-1	0
6	$-\sqrt{2}/2$	$-\sqrt{2}/2$
7	0	-1
8	$\sqrt{2}/2$	$-\sqrt{2}/2$
9	1	0

и

$$x_2 = x_1 \cos \delta\theta - y_1 \sin \delta\theta = (1)(\sqrt{2}/2) - 0(\sqrt{2}/2) = (\sqrt{2}/2),$$

$$y_2 = x_1 \sin \delta\theta + y_1 \cos \delta\theta = (1)(\sqrt{2}/2) + 0(\sqrt{2}/2) = (\sqrt{2}/2).$$

Результаты для остальных точек собраны в таблице 4-1.

Вспоминая двумерные преобразования из гл. 2, определим 3×3 преобразование, которое масштабирует полученные результаты с коэффициентом 2 и переносит центр окружности в точку $(2, 2)$:

$$[T] = [S][Tr_{xy}] = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 2 & 2 & 1 \end{bmatrix}.$$

Применяя это преобразование к точке (x, y) , получаем

$$\begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} [T] = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 2 & 2 & 1 \end{bmatrix} = [4 \quad 2 \quad 1],$$

как и ожидалось. Полный результат приведен в таблице 4-2 и на рис. 4-5.

Таблица 4-2 Результаты для окружности радиуса 2 с центром в $(2, 2)$

i	x_i	y_i
1	4	2
2	3.414	3.414
3	2	4
4	0.586	3.414
5	0	2
6	0.586	0.586
7	2	0
8	3.414	0.586
9	4	2

Ограничав диапазон параметра θ , можно получить дуги окружности. Например, $0 \leq \theta \leq \pi/2$ соответствует четверти круга в первом квадранте, а $\pi \leq \theta \leq 3\pi/2$ — четверти круга в третьем квадранте.

4-6 ПАРАМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ЭЛЛИПСА

Достаточно хорошее представление окружности получается, если соединить отрезками некоторое количество равномерно распределенных на ней точек. Однако, если рассчитывать точки эллипса через равные приращения угла, изображение будет неверным, как показано штриховой линией на рис. 4-6. Особенно сильно неточности проявляются на концах, где кривизна слишком велика и требуется большее количество точек.

Другой метод основан на равных приращениях по периметру и дает хороший результат для достаточно большого количества точек. Недостатки его в том, что указывается слишком много точек на сторонах с малой кривизной, и вычисление равных частей периметра требует сложного расчета эллиптического интеграла. Что нам необходимо, это малые приращения параметра θ на концах, где кривизна велика, и большие приращения параметра θ вдоль сторон с малой кривизной.

Такое распределение точек получается из параметрического представления эллипса с центром в начале координат, большой полуосью a и малой полуосью b :

$$\begin{aligned} x &= a \cos \theta, \\ y &= b \sin \theta, \end{aligned} \quad (4-6)$$

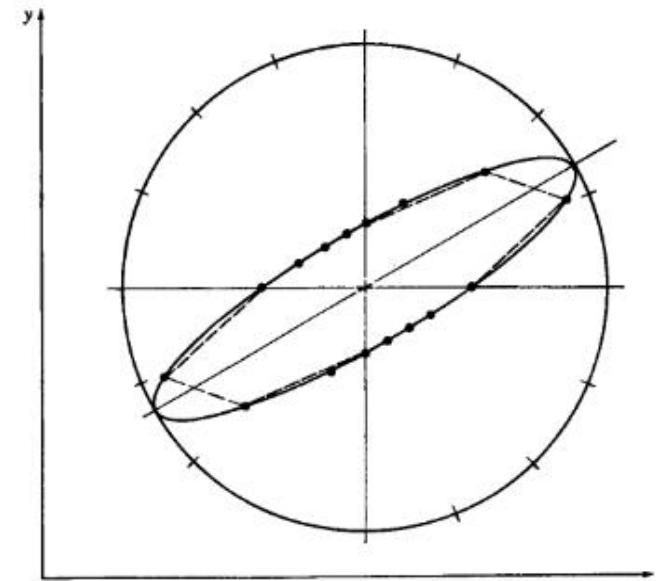


Рис. 4-6 Равноугольное представление сильно вытянутого эллипса.

где θ — параметр. Диапазон от 0 до 2π задает полный эллипс. Требуемое распределение точек порождается равномерными приращениями параметра θ .

Рассматривая производные x и y :

$$\begin{aligned} dx &= -a \sin \theta d\theta, \\ dy &= b \cos \theta d\theta, \end{aligned} \quad (4-7)$$

получаем, что при θ , близком к 0 или π , т. е. у концов, $|dx| \approx 0$ и $|dy| \approx bd\theta$. Если θ близко к $\pi/2$ или $3\pi/2$, т. е. вдоль сторон, то $|dx| \approx ad\theta$ и $|dy| \approx 0$. Таким образом, около концов, где кривизна более высокая, точки располагаются чаще, а вдоль сторон, где кривизна меньше, — реже. Отношение приращений периметра концов к приращениям вдоль сторон приблизительно равно b/a . Отметим, что для окружности ($b = a$) достигается оптимальное представление — равные приращения параметра или угла.

Если задано фиксированное количество точек на эллипсе, можно, пользуясь формулами суммы углов, получить эффективный алгоритм [4-2]. Координаты любой точки на эллипсе:

$$\begin{aligned} x_{i+1} &= a \cos(\theta_i + \delta\theta), \\ y_{i+1} &= b \sin(\theta_i + \delta\theta), \end{aligned}$$

где $\delta\theta = 2\pi/(n-1)$ — фиксированное приращение θ , n — количество точек на периметре, θ_i — значение параметра для точки x_i , y_i .

По формуле суммы углов

$$\begin{aligned} x_{i+1} &= a(\cos \theta_i \cos \delta\theta - \sin \theta_i \sin \delta\theta), \\ y_{i+1} &= b(\cos \theta_i \sin \delta\theta + \cos \delta\theta \sin \theta_i). \end{aligned}$$

Применяя уравнение (4-4) с $\theta = \theta_i$, перепишем эти уравнения:

$$\begin{aligned} x_{i+1} &= x_i \cos \delta\theta - \left(\frac{a}{b}\right) y_i \sin \delta\theta, \\ y_{i+1} &= \left(\frac{b}{a}\right) x_i \sin \delta\theta + y_i \cos \delta\theta, \end{aligned} \quad (4-8)$$

Так как $\delta\theta$, a и b — константы, полученный алгоритм содержит во внутреннем цикле только четыре умножения, одно сложение и вычитание и достаточно эффективен. В работе [4-2] показано, что алгоритм дает многоугольник максимальной площади, вписанный в эллипс. Результат представлен на рис. 4-7.

Чтобы получить эллипс с центром не в начале координат и с главной осью, расположенной под углом к горизонтали, его поворачивают вокруг начала координат, а затем переносят.

Пример 4-4 Параметрическое построение эллипса

Построить эллипс с большой полуосью $a = 4$ и малой полуосью $b = 1$, под углом 30° к горизонтали, с центром в точке $(2, 2)$.

Сначала построим эллипс с центром в начале координат с помощью 32 точек ($n = 33$, так как первая и последняя точки совпадают). Мы рассмотрим только точки первого квадранта, т. е. диапазон параметра от 0 до $\pi/2$.

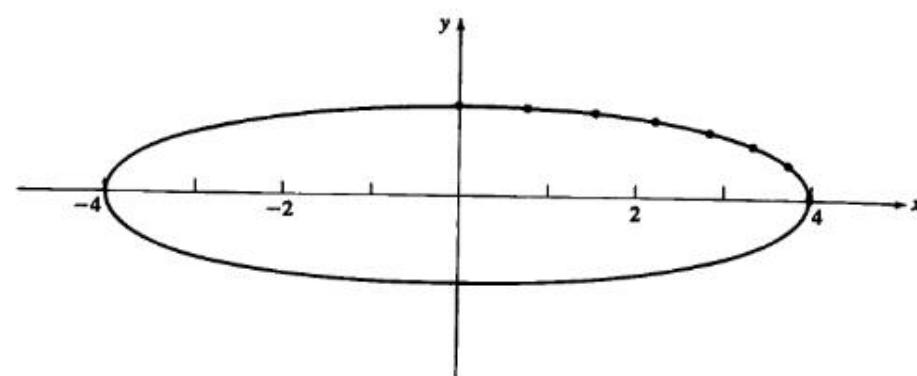


Рис. 4-7 Параметрическое представление эллипса.

Приращение параметра

$$\delta\theta = \frac{2\pi}{(n-1)} = \frac{2\pi}{32} = \frac{\pi}{16}.$$

Пользуясь уравнением (4-6) с начальным значением $\theta = 0$, найдем x и y :

$$\begin{aligned} x_1 &= a \cos \theta_1 = (4) \cos(0) = 4, \\ y_1 &= b \sin \theta_1 = (1) \sin(0) = 0. \end{aligned}$$

Вычислим величины $a/b = 4$, $b/a = 1/4$ и

$$\begin{aligned} \sin \delta\theta &= \sin \frac{\pi}{16} = 0.195, \\ \cos \delta\theta &= \cos \frac{\pi}{16} = 0.981. \end{aligned}$$

Теперь по формуле (4-8) найдем вторую точку

$$\begin{aligned} x_2 &= x_1 \cos \delta\theta - \left(\frac{a}{b}\right) y_1 \sin \delta\theta = (4)(0.981) - (4)(0)(0.195) = 3.92, \\ y_2 &= \left(\frac{b}{a}\right) x_1 \sin \delta\theta + y_1 \cos \delta\theta = (1/4)(4)(0.195) + (0)(0.981) = 0.195. \end{aligned}$$

Результаты для оставшихся точек первого квадранта представлены в таблице 4-3 и на рис. 4-7.

Нам требуется комбинированное двумерное преобразование (гл. 2) поворота на 30° вокруг центра координат и переноса центра в точку $(2, 2)$.

Характерно, что

$$\begin{aligned} [T] &= [R][Tr_{xy}] = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0.866 & 0.5 & 0 \\ -0.5 & 0.866 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 0.866 & 0.5 & 0 \\ -0.5 & 0.866 & 0 \\ 2 & 2 & 1 \end{bmatrix}. \end{aligned}$$

Таблица 4-3 Точки эллипса с центром в начале координат для первого квадранта

i	x_i	y_i
1	4.0	0
2	3.923	0.195
3	3.696	0.383
4	3.326	0.556
5	2.828	0.707
6	2.222	0.831
7	1.531	0.924
8	0.780	0.981
9	0	1.0

Применим это преобразование для точек (x_1, y_1) и (x_2, y_2) :

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix} [T] = \begin{bmatrix} 4 & 0 & 1 \\ 3.923 & 0.195 & 1 \end{bmatrix} \begin{bmatrix} 0.866 & 0.5 & 0 \\ -0.5 & 0.866 & 0 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 5.464 & 4 & 1 \\ 5.3 & 4.131 & 1 \end{bmatrix}.$$

Результаты приведены в табл. 4-4 и на рис. 4-8.

Таблица 4-4 Повернутый и перенесенный эллипс

i	x_i	y_i
1	5.464	4.0
2	5.3	4.131
3	5.009	4.179
4	4.603	4.144
5	4.096	4.027
6	3.509	3.831
7	3.864	3.565
8	2.185	3.240
9	1.5	2.866

4-7 ПАРАМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ПАРАБОЛЫ

Рассмотрим параболу с вершиной в центре координат, раскрыв вправо, т. е. с осью симметрии — положительной полуосью x . На рис. 4-9 изображена верхняя ветвь такой параболы. В прямоугольных координатах непараметрическое представление параболы:

$$y^2 = 4ax.$$

Параметрическое представление имеет вид

$$x = t \cos^2 \phi,$$

$$y = \pm 2\sqrt{a \cos \phi},$$

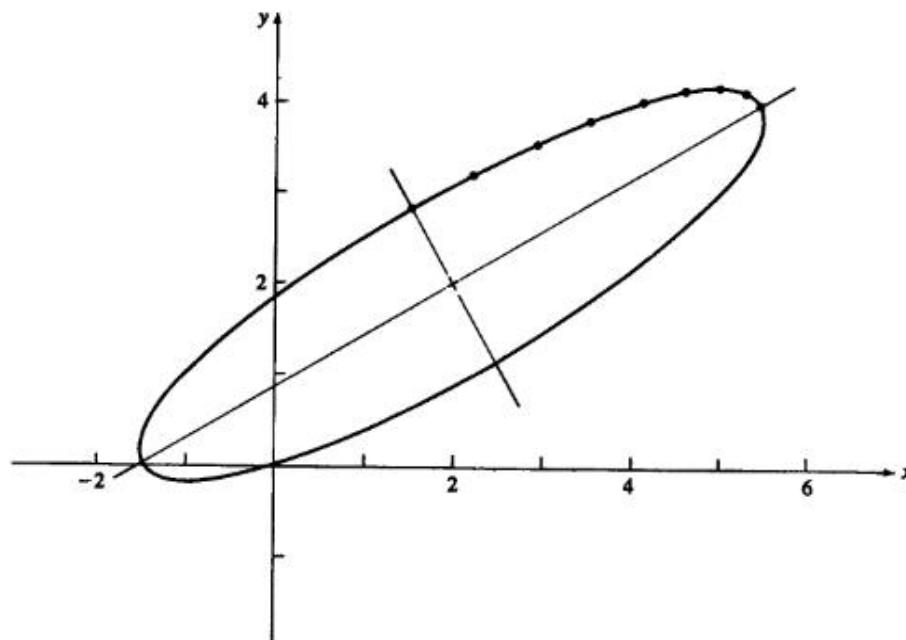


Рис. 4-8 Эллипс после поворота и переноса.

где $0 \leq \phi \leq \pi/2$. Хотя оно обеспечивает достаточно хорошее изображение, Смит [4-2] указал, что получаемая фигура не является фигурой с максимальной вписанной площадью и поэтому это не оптимальный вариант. Другое параметрическое представление действительно дает фигуру с наибольшей вписанной площадью:

$$\begin{aligned} x &= a\theta^2, \\ y &= 2a\theta, \end{aligned} \quad (4-9)$$

где $0 \leq \theta \leq \infty$ соответствует всей верхней ветви параболы. В отличие от эллипса парабола не замкнутая кривая, поэтому изображаемая часть должна быть ограничена.

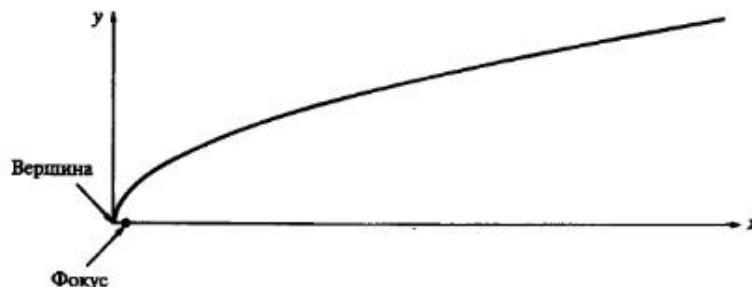


Рис. 4-9 Парабола.

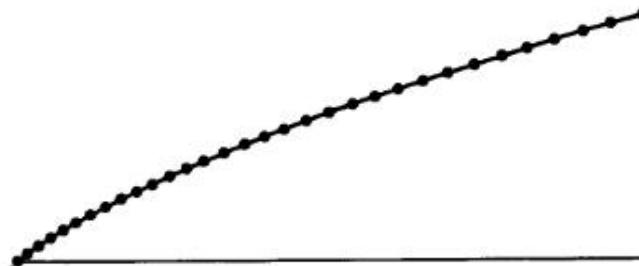


Рис. 4-10 Параметрически сгенерированная парабола.

ничена минимальным и максимальным значением параметра. Это можно сделать несколькими способами. Если диапазон изменения координаты x ограничен, то

$$\theta_{\min} = \sqrt{\frac{x_{\min}}{a}}, \quad \theta_{\max} = \sqrt{\frac{x_{\max}}{a}}. \quad (4-10)$$

Если ограничен диапазон изменения y , то

$$\theta_{\min} = \frac{y_{\min}}{2a}, \quad \theta_{\max} = \frac{y_{\max}}{2a}. \quad (4-11)$$

Установив θ_{\min} и/или θ_{\max} , можно построить параболу в первом квадранте. Параболы в других квадрантах, со смещенным центром и в других ориентациях строятся с помощью отражения, поворота и переноса.

Параболу можно построить также, пользуясь приращениями параметра. Пусть на параболе задано фиксированное количество точек, т. е. приращение параметра θ постоянно. Для $\theta_{i+1} = \theta_i + \delta\theta$ уравнение (4-9) принимает вид

$$\begin{aligned} x_{i+1} &= a\theta_i^2 + 2a\theta_i\delta\theta + a(\delta\theta)^2, \\ y_{i+1} &= 2a\theta_i + 2a\delta\theta. \end{aligned}$$

Используя уравнение (4-9) с $\theta = \theta_i$, перепишем формулы

$$\begin{aligned} x_{i+1} &= x_i + y_i\delta\theta + a(\delta\theta)^2, \\ y_{i+1} &= y_i + 2a\delta\theta. \end{aligned} \quad (4-12)$$

Расчет очередной точки требует трех сложений и одного умножения во внутреннем цикле алгоритма. На рис. 4-10 приведен пример параболы, сгенерированной по рекурсивным формулам (4-12).

Пример 4-5 Параметрическое представление параболы

Построить параболический сегмент в первом квадранте при $1 \leq x \leq 4$ для параболы

$$x = a\theta^2 = \theta^2, \quad y = 2a\theta = 2\theta$$

при $a = 1$.

Сначала найдем границы θ . В соответствии с уравнением (4-10) θ_{\min} и θ_{\max} определяются так:

$$\theta_{\min} = \sqrt{\frac{x_{\min}}{a}} = \sqrt{\frac{1}{1}} = 1,$$

$$\theta_{\max} = \sqrt{\frac{x_{\max}}{a}} = \sqrt{\frac{4}{1}} = 2.$$

Пусть на сегменте расположено 10 точек, тогда

$$\delta\theta = \frac{\theta_{\max} - \theta_{\min}}{n-1} = \frac{2-1}{10-1} = \frac{1}{9}.$$

Начиная с $\theta_1 = \theta_{\min}$, $x_1 = 1$, из уравнения (4-9) получаем

$$y_1 = 2a\theta_1 = (2)(1) = 2.$$

Из уравнения (4-12)

$$\begin{aligned} x_2 &= x_1 + y_1\delta\theta + (\delta\theta)^2 = 1 + (2)\left(\frac{1}{9}\right) + \left(\frac{1}{9}\right)^2 = 1.235, \\ y_2 &= y_1 + 2\delta\theta = 2 + (2)\left(\frac{1}{9}\right) = 2.222. \end{aligned}$$

Окончательный результат приведен в табл. 4-5 и на рис. 4-11.

Таблица 4-5 Результаты для сегмента параболы

i	x_i	y_i
1	1.0	2.0
2	1.235	2.222
3	1.494	2.444
4	1.778	2.667
5	2.086	2.889
6	2.420	3.111
7	2.778	3.333
8	3.160	3.556
9	3.568	3.778
10	4.0	4.0

В некоторых случаях более удобны другие параметрические представления, вид которых зависит от прикладной задачи и имеющихся данных. Например, если надо нарисовать дугу параболы между двумя точками и учитывать наклон, то предлагается следующее представление:

$$\begin{aligned} x(t) &= (Q_x - 2R_x + R_x)t^2 + 2(R_x - P_x)t + P_x, \\ y(t) &= (Q_y - 2R_y + P_y)t^2 + 2(R_y - P_y)t + P_y, \quad 0 \leq t \leq 1.0, \end{aligned} \quad (4-13)$$

где t — параметр, а две конечные точки $P = [P_x \quad P_y]$ и $Q = [Q_x \quad Q_y]$. Точка $R = [R_x \quad R_y]$ это точка пересечения касательных в конечных точках. Три вершины P , Q , R определяют параболу, как показано на рис. 4-12. Более общий метод построения кривых с помощью вершин незамкнутого многоугольника был разработан Безье и рассматривается в следующей главе.

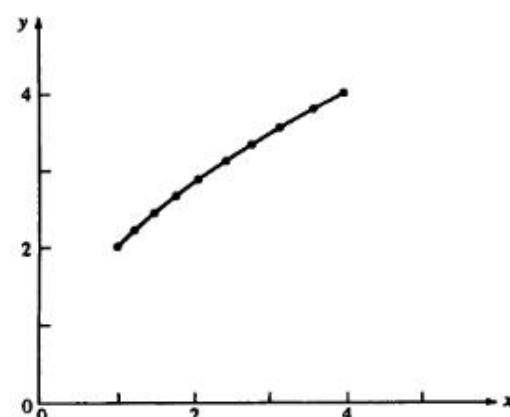


Рис. 4-11 Сегмент параболы.

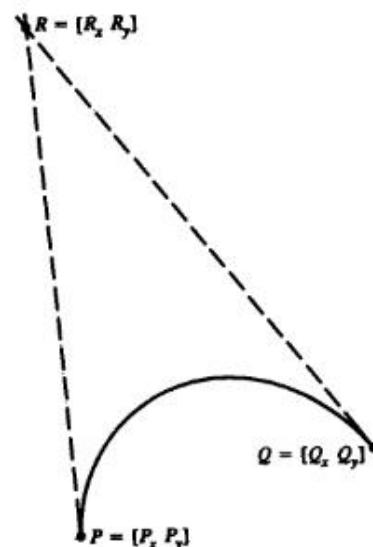


Рис. 4-12 Задание параметрической параболы вершинами.

4-8 ПАРАМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ГИПЕРБОЛЫ

Построим гиперболу с центром в начале координат и осью симметрии, совпадающей с осью x . Ее непараметрическое представление в прямоугольных координатах:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

При этом вершина находится в точке $(a, 0)$, оси асимптот $\pm b/a$. Вид параметрического представления:

$$\begin{aligned} x &= \pm a \sec \theta, \\ y &= \pm b \operatorname{tg} \theta, \end{aligned} \quad (4-14)$$

где $0 \leq \theta \leq \pi/2$, дает искомую гиперболу. Смит [4-2] указывает, что для такого представления площадь вписанного многоугольника не максимальна. Однако она близка к максимальной, и с помощью формулы суммы углов можно получить эффективный алгоритм. Вспомним, что

$$\sec(\theta + \delta\theta) = \frac{1}{\cos(\theta + \delta\theta)} = \frac{1}{\cos \theta \cos \delta\theta - \sin \theta \sin \delta\theta}$$

и

$$\operatorname{tg}(\theta + \delta\theta) = \frac{\operatorname{tg} \theta + \operatorname{tg} \delta\theta}{1 - \operatorname{tg} \theta \operatorname{tg} \delta\theta}.$$

Подставим в уравнения (4-14)

$$\begin{aligned} x_{i+1} &= \pm a \sec(\theta + \delta\theta) = \pm \frac{ab/\cos \theta}{b \cos \delta\theta - b \operatorname{tg} \theta \sin \delta\theta}, \\ y_{i+1} &= \pm b \operatorname{tg}(\theta + \delta\theta) = \pm \frac{b \operatorname{tg} \theta + b \operatorname{tg} \delta\theta}{1 - \operatorname{tg} \theta \operatorname{tg} \delta\theta}. \end{aligned}$$

Используя уравнения (4-14) с $\theta = \theta_i$, перепишем эти уравнения как

$$\begin{aligned} x_{i+1} &= \pm \frac{bx_i}{b \cos \delta\theta - y_i \sin \delta\theta}, \\ y_{i+1} &= \pm \frac{b(y_i + b \operatorname{tg} \delta\theta)}{b - y_i \operatorname{tg} \delta\theta}. \end{aligned} \quad (4-15)$$

Другое параметрическое представление гиперболы, дающее максимальную вписанную площадь:

$$\begin{aligned} x &= a \operatorname{ch} \theta, \\ y &= b \operatorname{sh} \theta. \end{aligned} \quad (4-16)$$

Гиперболические функции определяются как $\operatorname{ch} \theta = (e^\theta + e^{-\theta})/2$ и $\operatorname{sh} \theta = (e^\theta - e^{-\theta})/2$. При изменении от 0 до бесконечности проходит вся гипербола. Формула для суммы углов для и

$$\begin{aligned} \operatorname{ch}(\theta + \delta\theta) &= \operatorname{ch} \theta \operatorname{ch} \delta\theta + \operatorname{sh} \theta \operatorname{sh} \delta\theta, \\ \operatorname{sh}(\theta + \delta\theta) &= \operatorname{sh} \theta \operatorname{ch} \delta\theta + \operatorname{ch} \theta \operatorname{sh} \delta\theta. \end{aligned}$$

Это позволяет записать уравнения (4-16) как

$$\begin{aligned} x_{i+1} &= a(\operatorname{ch} \theta \operatorname{ch} \delta\theta + \operatorname{sh} \theta \operatorname{sh} \delta\theta), \\ y_{i+1} &= b(\operatorname{sh} \theta \operatorname{ch} \delta\theta + \operatorname{ch} \theta \operatorname{sh} \delta\theta) \end{aligned}$$

или

$$\begin{aligned}x_{i+1} &= x_i \operatorname{ch} \delta\theta + \left(\frac{a}{b}\right) y_i \operatorname{sh} \delta\theta, \\y_{i+1} &= \left(\frac{a}{b}\right) x_i \operatorname{sh} \delta\theta + y_i \operatorname{ch} \delta\theta.\end{aligned}\quad (4-17)$$

Чтобы ограничить область гиперболы, необходимо установить минимальное и максимальное значения. Пусть ветвь гиперболы лежит в первом и четвертом квадранте и рассматривается часть при $x_{\min} \leq x \leq x_{\max}$. Тогда

$$\begin{aligned}\theta_{\min} &= \operatorname{Arch} \left(\frac{x_{\min}}{a} \right), \\ \theta_{\max} &= \operatorname{Arch} \left(\frac{x_{\max}}{a} \right),\end{aligned}\quad (4-18)$$

где обратный гиперболический косинус получен как

$$\operatorname{Arch} x = \ln(x + \sqrt{x^2 - 1}). \quad (4-19)$$

Остальные границы определяются аналогично. Пример части гиперболы в первом квадранте, полученной этим методом, показан на рис. 4-13.

Пример 4-6 Параметрическая гипербола

С помощью параметрического представления (4-16) найти восемь точек на сегменте гиперболы в первом квадранте при $a = 2$, $b = 1$, для $4 \leq x \leq 8$.

Сначала определим границы параметра из уравнений (4-18) и (4-19)

$$\begin{aligned}\theta_{\max} &= \operatorname{Arch} \left(\frac{x_{\max}}{a} \right) = \ln \left(\frac{x_{\max}}{a} + \sqrt{\left(\frac{x_{\max}}{a} \right)^2 - 1} \right) = \\&= \ln(4 + \sqrt{16 - 1}) \\&= 2.063.\end{aligned}$$

Аналогично,

$$\theta_{\min} = 1.317.$$

Итак,

$$\delta\theta = \frac{\theta_{\max} - \theta_{\min}}{n-1} = \frac{2.063 - 1.317}{7} = 0.107$$

и

$$\begin{aligned}\operatorname{sh}(\delta\theta) &= \operatorname{sh}(0.107) = \\&= \frac{e^{0.107} - e^{-0.107}}{2} = \\&= 0.107, \\ \operatorname{ch}(\delta\theta) &= \operatorname{ch}(0.107) = \\&= \frac{e^{0.107} + e^{-0.107}}{2} = \\&= 1.006.\end{aligned}$$

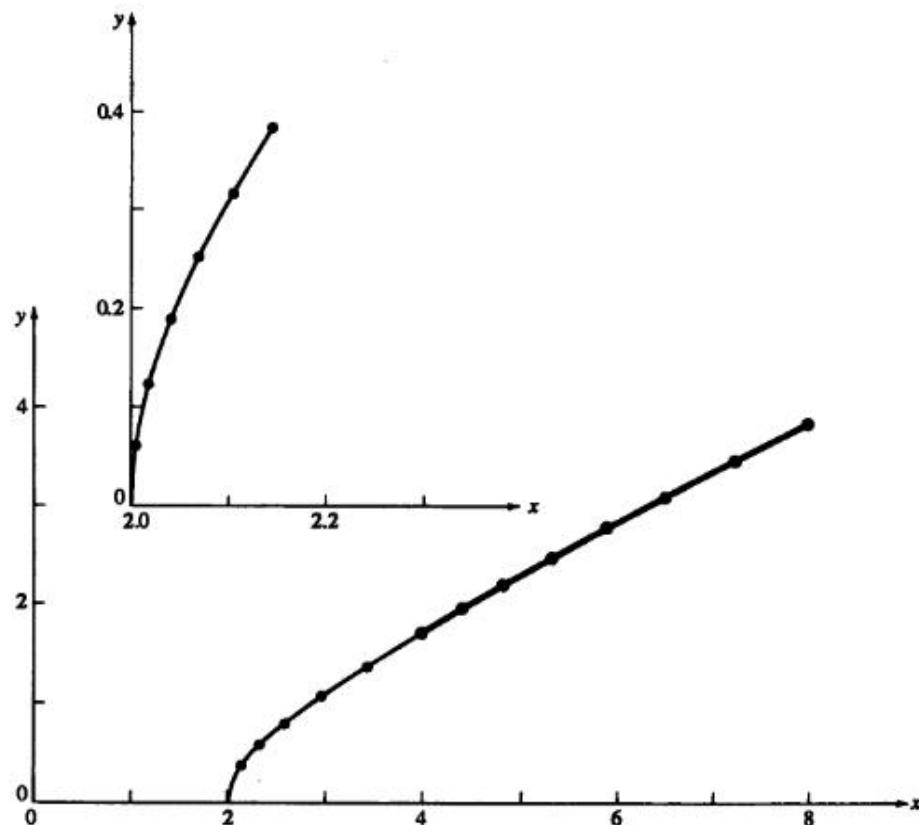


Рис. 4-13 Параметрическая гипербола.

Уравнения (4-16) с $\theta_1 = \theta_{\min}$ дают

$$\begin{aligned}x_1 &= a \operatorname{ch}(\theta_{\min}) = \\&= (2) \operatorname{ch}(1.317) = \\&= 4.000, \\y_1 &= b \operatorname{sh}(\theta_{\min}) = \\&= (1) \operatorname{sh}(1.317) = \\&= 1.732.\end{aligned}$$

Затем из уравнений (4-17)

$$\begin{aligned}x_2 &= x_1 \operatorname{ch}(\delta\theta) + (a/b)y_1 \operatorname{sh}(\delta\theta) \\&= (4)(1.006) + (2)(1.732)(0.107) \\&= 4.393, \\y_2 &= (b/a)x_1 \operatorname{sh}(\delta\theta) + y_1 \operatorname{ch}(\delta\theta) \\&= (1/2)(4)(0.107) + (1.732)(1.006) = 1.956.\end{aligned}$$

Результаты вычислений приведены в табл. 4-6 и выделены жирной линией на рис. 4-13.

Таблица 4-6 Результаты для сегмента гиперболы

<i>i</i>	<i>x_i</i>	<i>y_i</i>
1	4	1.732
2	4.393	1.956
3	4.836	2.201
4	5.334	2.472
5	5.892	2.771
6	6.518	3.102
7	7.218	3.468
8	8	3.873

4-9 ПРОЦЕДУРА ИСПОЛЬЗОВАНИЯ КОНИЧЕСКИХ СЕЧЕНИЙ

Плоские геометрические фигуры часто используются в прикладной машинной графике, например для разработки деталей или в чертежных системах. Особенно важны здесь проблемы определения положения конического сечения и его пересечения и/или точек касания с другим коническим сечением или прямой. На первый взгляд задача несложная, но для произвольно расположенных на плоскости элементов могут существовать различные решения. Выбор правильного решения не всегда очевиден, что приводит к нелинейным расчетам. Ниже приводятся методы, исключающие нелинейную математику и упрощающие линейные результаты. Кроме того, объясняется, как выбрать необходимое решение, если оно неоднозначно.

Основная идея заключается в том, чтобы методами машинной графики, в частности плоским поворотом и переносом, перевести расчеты в первый квадрант в стандартной конфигурации. Если фигура включает коническое сечение, его центр (для параболы и гиперболы — вершина) располагается в начале координат. В общем случае неизвестные центры, вершины, точки касания и пересечения вычисляются с помощью непараметрических уравнений, а параметрическое представление используется для изображения фигур. Разделение расчета и построения позволяет максимально использовать преимущества обоих представлений.

Рассмотрим подробный пример, чтобы лучше представить себе суть и возможности этого метода. Одна из простейших и часто используемых задач — построить окружность по трем точкам. Чертеж приведен на рис. 4-14, где $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $P_3(x_3, y_3)$ заданные точки. Необходимо определить центр и радиус окружности. Для этого нужно решить систему из трех нелинейных уравнений с переменными h , k , координатами центра (x, y) и радиусом R :

$$(x_1 - h)^2 + (y_1 - k)^2 = R^2, \quad (4-20a)$$

$$(x_2 - h)^2 + (y_2 - k)^2 = R^2, \quad (4-20b)$$

$$(x_3 - h)^2 + (y_3 - k)^2 = R^2. \quad (4-20c)$$

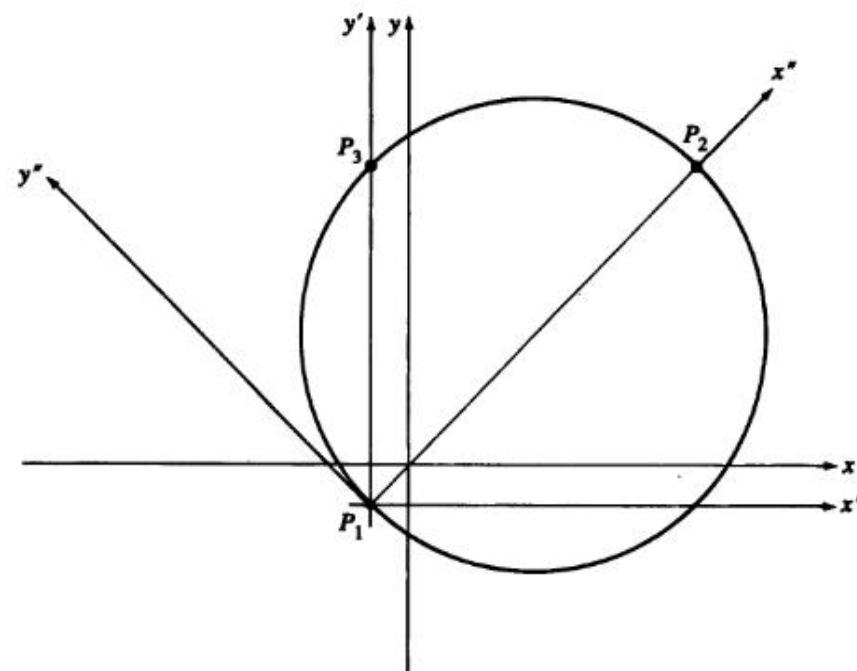


Рис. 4-14 Окружность, проходящая через три заданные точки.

Вычитание уравнения (4-20b) из уравнений (4-20a) и (4-20c) устраниет нелинейность. Собственно алгоритм решения — это $[(a)-(b)](x_3-x_2)-[(c)-(b)](x_1-x_2)$, где буквы в скобках взяты из уравнений (4-20). Уравнение, выраженное через неизвестную, таково:

$$k = \frac{[(x_1^2 - x_2^2) + (y_1^2 - y_2^2)](x_3 - x_2)}{2[(y_1 - y_2)(x_3 - x_2) - (y_3 - y_2)(x_1 - x_2)]} - \frac{[(x_3^2 - x_2^2) + (y_3^2 - y_2^2)](x_1 - x_2)}{2[(y_1 - y_2)(x_3 - x_2) - (y_3 - y_2)(x_1 - x_2)].} \quad (4-21)$$

Аналогично из $[(a) - (b)]$ и уравнения (4-21) вычисляется

$$h = \frac{(x_1^2 - x_2^2) + (y_1^2 - y_2^2) - 2k(y_1 - y_2)}{2(x_1 - x_2)}, \quad (4-22)$$

R находится из любого из уравнений (4-20).

Уравнения (4-21) и (4-22) содержат ряд трудностей. В частности, если знаменатель дроби равен нулю, требуется другие методы решения. Также необходимы условия проверки на бесконечный радиус и коллинеарные точки.

Уравнения для h и k можно упростить переносом начала координат в точку (см. рис. 4-14). Тогда

$$h'^2 + k'^2 = R^2, \quad (4-23a)$$

$$(x'_2 - h')^2 + (y'_2 - k')^2 = R^2, \quad (4-23b)$$

$$(x'_3 - h')^2 + (y'_3 - k')^2 = R^2. \quad (4-23c)$$

Вычитая (a) из (b) и (c), получаем систему из двух линейных уравнений. Алгоритм решения: $[(b) - (a)]x'_3 - [(c) - (a)]x'_2$, что дает

$$k' = \frac{x'_3 x'_2 (x'_2 - x'_3) + y'^2_2 x'_3 - y'^2_3 x'_2}{2(y'_2 x'_3 - y'_3 x'_2)}, \quad (4-24)$$

$$h' = \frac{x'^2_2 + y'^2_2 - 2y'_2 k'}{2x'_2}, \quad (4-25)$$

$$R = (h'^2 + k'^2)^{\frac{1}{2}}. \quad (4-26)$$

Преобразование h' , k' обратно к первоначальной системе координат дает h , k . При этом знаменатель уравнений (4-24) или (4-25) опять может обратиться в нуль, и требуются условия проверки на бесконечный радиус и коллинеарные точки. Эти трудности преодолеваются с помощью следующей процедуры:

Перенести начало координат в одну из точек.

Повернуть оси координат, чтобы одна из оставшихся точек лежала на положительной полуоси x .

Проверить коллинеарность точек.

Найти центр и радиус окружности в перенесенной и повернутой системе координат.

Повернуть обратно к ориентации промежуточной системы координат.

Перенести начало координат в первоначальное положение и найти центр окружности в исходной системе координат.

На рис. 4-14 начало координат перенесено в точку P_1 , а положительная полуось x'' проходит через точку P_2 . Уравнения в координатах (x'', y'') :

$$h''^2 + k''^2 = R^2, \quad (4-27a)$$

$$(x''_2 - h'')^2 + k''^2 = R^2, \quad (4-27b)$$

$$(x''_3 - h'')^2 + (y''_3 - k'')^2 = R^2. \quad (4-27c)$$

Вычтем уравнение (4-27a) из (4-27b)

$$x''_2^2 - 2x''_2 h'' = 0,$$

или

$$h'' = x''_2 / 2. \quad (4-28)$$

Вычтем уравнение (4-27a) из (4-27c) и подставим уравнение (4-28):

$$k'' = \frac{x''_3 (x''_3 - x''_2) + y''_3}{2y''_3}, \quad (4-29)$$

а из уравнения (4-27a) получим

$$R = (h''^2 + k''^2)^{\frac{1}{2}}. \quad (4-30)$$

Уравнение (4-29) показывает, что при $y''_3 = 0$ значение k бесконечно, однако это случается только тогда, когда все три точки лежат на одной прямой. Проверка этого условия очень проста.

Провести окружность через три точки достаточно легко. Здесь не требуется выбора из множества решений, как в большинстве геометрических конструкций, включающих конические сечения. Метод выбора предложен в работе Роджерса [4-3].

Роджерс рассматривает левое, правое, ближнее, дальнее, внутреннее и внешнее решение. Левое и правое существуют относительно направления от одной известной точки к другой; ближнее и дальнее — относительно линии пересечения; внутреннее и внешнее — относительно окружности.

Эти понятия справедливы как в ориентации пользователя, т. е. в любом месте пространства, так и в стандартной конфигурации в первом квадранте. Простой пример для иллюстрации понятий левое—правое и ближнее—дальнее это определение центра и точек касания окружности известного радиуса R , касающейся двух пересекающихся прямых P и L . На практике задача возникает, например, при изготовлении закругленных углов. Для любого радиуса всегда существует четыре решения. Стратегия выбора такова:

Найти точку пересечения прямых. Если решения нет, прямые параллельны.

Перенести начало координат в точку пересечения.

Повернуть вокруг нового центра координат, чтобы одна из линий совпала с осью x .

Найти угол между другой линией и осью x .

Повернуть вокруг нового начала координат, чтобы положительная полуось x совпала с биссектрисой пересекающихся прямых.

Найти центр и точки касания.

Перевести в первоначальную систему координат.

На рис. 4-15 изображен этот случай в стандартной конфигурации. Две пересекающиеся линии P и L заданы точками P_1, P_2, L_1, L_2 соответственно. Для каждого возможного решения одна из координат центра (h, k) лежит на оси координат. Чтобы выбрать нужное решение, требуется указать, находится ли оно слева или справа от линии P , если смотреть от P_1 и P_2 , и ближнее оно или дальнее относительно линии L с той же точки зрения. На рис. 4-15 C_1 — правое и дальнее, а C_3 — левое и ближнее. Центр

$$h = R / \sin \theta,$$

$$k = 0,$$

где θ — острый угол между положительной осью x и P . Точки касания с линиями P и L

$$x_t = h - R \sin \theta,$$

$$y_t = \pm R \cos \theta.$$

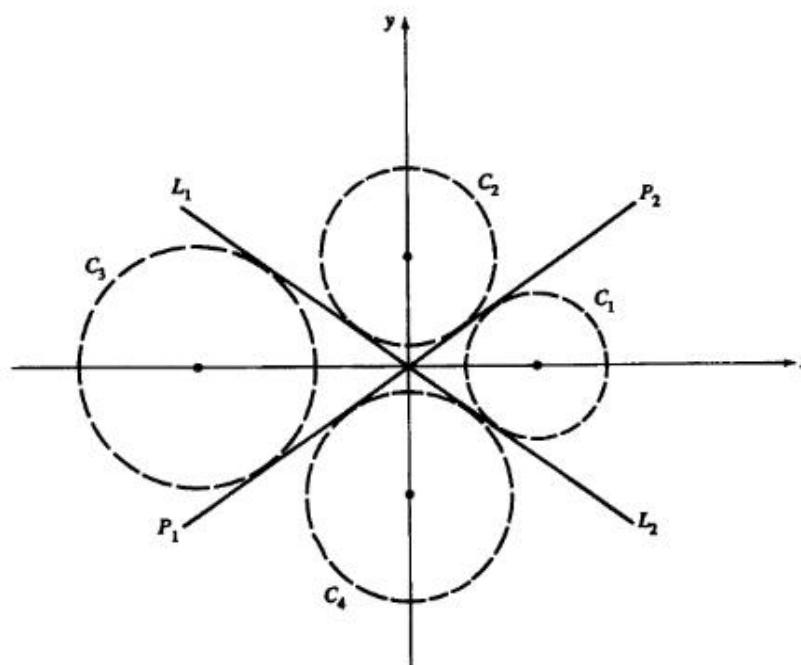


Рис. 4-15 Окружности, касающиеся двух пересекающихся прямых.

Таблица 4-7 Окружность, касающаяся двух пересекающихся линий

Ориентация относительно линии \$P\$	Ориентация относительно линии \$L\$	Центр	Точки касания
Левая	Дальняя	\$h = 0\$	\$x_t = \pm R \cos(90 - \theta) = \pm R \sin(\theta)\$
		\$k = R / \sin(90 - \theta) = R / \cos(\theta)\$	\$y_t = k - R \cos(\theta)\$
Левая	Ближняя	\$h = -R / \sin(\theta)\$	\$x_t = h + R \sin(\theta)\$
		\$k = 0\$	\$y_t = \pm R \cos(\theta)\$
Правая	Дальняя	\$h = -R / \sin(\theta)\$	\$x_t = h - R \sin(\theta)\$
		\$k = 0\$	\$y_t = \pm R \cos(\theta)\$
Правая	Ближняя	\$h = 0\$	\$x_t = \pm R \cos(90 - \theta) = \pm R \sin(\theta)\$
		\$k = -R / \sin(90 - \theta) = -R / \cos(\theta)\$	\$y_t = k + R \cos(\theta)\$

Полный перечень приведен в табл. 4-7. Более сложные примеры с использованием понятий внутреннее-внешнее содержатся в работе [4-3]. Возможно расширение для других конических сечений. В работе [4-4] рассматривается векторный подход к решению подобных проблем.

4-10 ОБЩИЕ УРАВНЕНИЯ КОНИЧЕСКИХ СЕЧЕНИЙ

Все конические сечения из предыдущих разделов являются частными случаями кривых, задаваемых общим уравнением второго порядка

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (4-31)$$

где \$A, B, C, D, E, F\$ — константы¹. Коническое сечение — это любая плоская кривая, удовлетворяющая уравнению (4-31). Уравнение интересно не только само по себе, но и для последующего обсуждения рациональных конических сечений и квадратичных поверхностей. Для простоты и полноты будем пользоваться методами линейной алгебры.

Конические сечения являются центральными — эллипс и гипербола (окружность это частный случай эллипса) или нецентральными — парабола. Кроме того, существует ряд вырожденных форм, которые все центральны. Необходимо определить для заданных значений констант, какое сечение задает уравнение (4-31) — центральное или нецентральное. Также нужно выделить все вырожденные случаи.

Заметим, что уравнение (4-31) можно записать в матричной форме

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} A & D/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0$$

или

$$[X][S][X]^T = 0. \quad (4-32)$$

Отметим также, что \$[S]\$ симметрична относительно главной диагонали.

Сначала приведем сечение к стандартному виду. Для центрального сечения (эллипс или гипербола) это значит, что центр находится в начале координат, а оси располагаются вдоль осей координат. Для нецентрального сечения (парабола) ось симметрии параболы совпадает с положительной осью \$x\$, вершина находится в начале координат, и парабола раскрывается направо. Сечение приводится к стандартному виду переносом и вращением.

Для центральных сечений линейные члены уравнений (4-31) или (4-32) уничтожаются переносом центра сечения в начало координат. После этого уравнение (4-32) принимает вид

$$[X][Tr][S][Tr]^T[X]^T = 0, \quad (4-33)$$

где матрица переноса \$[Tr]\$ такова:

$$[Tr] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & 1 \end{bmatrix}.$$

¹ Уравнение (4-31) — это модифицированная версия неявного уравнения из раздела 4-3 с \$A = a\$, \$B = 2b\$, \$C = c\$, \$D = 2d\$, \$E = 2\$, \$F = f\$.

После конкатенации матриц переноса и коэффициентов уравнение принимает вид

$$[X][S'][X]^T = 0, \quad (4-34)$$

где

$$[S'] = [Tr][S][Tr]^T = \left(\frac{1}{2}\right) \begin{bmatrix} 2A' & B' & D' \\ D' & 2C' & E' \\ D' & E' & 2F' \end{bmatrix} \quad (4-35)$$

и

$$A' = A$$

$$B' = B$$

$$C' = C$$

$$D' = 2Am + Bn + D$$

$$E' = Bm + 2Cn + E$$

$$2F' = mD' + nE' + (Dm + En + 2F).$$

Заметим, что $[S']$ также симметрична.

Следовательно, уравнение (4-31) преобразуется к

$$A'x^2 + B'xy + C'y^2 + D'x + E'y + F' = 0.$$

Коэффициенты переноса m и n для уничтожения линейных членов вычисляются из условия $D' = E' = 0$. Отсюда

$$2Am + Bn + D = 0,$$

$$Bm + 2Cn + E = 0$$

или в матричном виде

$$\begin{bmatrix} m & n \end{bmatrix} \begin{bmatrix} 2A & B \\ B & 2C \end{bmatrix} = \begin{bmatrix} -D \\ -E \end{bmatrix}, \quad (4-36)$$

что можно записать как $[M][L] = [Q]$.

Если $[L]$ инвертируема, существует решение для $[M]$, и сечение центрально, т.е. это эллипс или гипербола. Если $[L]$ не инвертируема, т.е. сингулярна, то решения для $[M]$ не существует, и сечение не центрально (парабола). Детерминант сингулярной матрицы равен нулю:

$$\det[L] = |L| = \begin{vmatrix} 2A & B \\ B & 2C \end{vmatrix} = 0 \quad (4-37)$$

или

$$B^2 - 4AC = 0.$$

Итак, уравнение (4-31) представляет параболу при $B^2 - 4AC = 0$ и центральное сечение при $B^2 - 4AC \neq 0$. Если сечение центрально и $B^2 - 4AC < 0$, уравнение представляет эллипс, а если $B^2 - 4AC > 0$ — гиперболу.

Независимо от инвертируемости $[L]$ оси сечения можно поворотом сделать параллельными осям координат. Вернемся к уравнению (4-32). Используем матрицу плоского поворота $[R]$

$$[X][R][S][R]^T[X]^T = 0, \quad (4-38)$$

где для угла поворота θ она такая:

$$[R] = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Конкатенация матриц дает

$$[X][S''][X]^T = 0, \quad (4-39)$$

где

$$[S''] = [R][S][R]^T = \left(\frac{1}{2}\right) \begin{bmatrix} 2A'' & B'' & D'' \\ B'' & 2C'' & E'' \\ D'' & E'' & 2F'' \end{bmatrix}$$

и

$$A'' = A \cos^2 \theta + B \cos \theta \sin \theta + C \sin^2 \theta,$$

$$B'' = 2(C - A) \cos \theta \sin \theta + B(\cos^2 \theta - \sin^2 \theta),$$

$$C'' = A \sin^2 \theta - B \cos \theta \sin \theta + C \cos^2 \theta,$$

$$D'' = D \cos \theta + E \sin \theta,$$

$$E'' = E \cos \theta - D \sin \theta,$$

$$F'' = F.$$

Опять заметим, что $[S'']$ симметрична. Если оси сечения параллельны осям координат, член Bxy в уравнении (4-31) отсутствует. Поэтому нулевой коэффициент дает угол поворота:

$$2(C - A) \cos \theta \sin \theta + B(\cos^2 \theta - \sin^2 \theta) = 0$$

или

$$(C - A) \sin 2\theta + B \cos 2\theta = 0.$$

Решим это уравнение относительно угла поворота θ :

$$\theta = \frac{1}{2} \operatorname{arctg} \left(\frac{B}{A - C} \right). \quad (4-40)$$

Для этого угла $[S'']$ принимает вид

$$[S''] = \left(\frac{1}{2}\right) \begin{bmatrix} 2A'' & 0 & D'' \\ 0 & 2C'' & E'' \\ D'' & E'' & 2F'' \end{bmatrix}_{\theta=\frac{1}{2}\arctg(\frac{B}{A-C})}$$

Уравнение (4-37) позволяет узнать, центрально ли сечение. Центральное сечение можно привести к стандартному виду комбинацией переноса и поворота:

$$[X][Tr][R][S][R]^T[Tr]^T[X]^T = 0. \quad (4-41)$$

Конкатенация внутренних матриц дает

$$\begin{aligned} [S'''] &= [Tr][R][S][R]^T[Tr]^T = \\ &= \left(\frac{1}{2}\right) \begin{bmatrix} 2A''' & B''' & D''' \\ B''' & 2C''' & E''' \\ D''' & E''' & 2F''' \end{bmatrix} = \\ &= \begin{bmatrix} A''' & 0 & \\ 0 & C''' & 0 \\ 0 & 0 & F''' \end{bmatrix} = \begin{bmatrix} \alpha & 0 & \\ 0 & \beta & 0 \\ 0 & 0 & -\kappa \end{bmatrix}, \end{aligned} \quad (4-42)$$

где

$$A''' = A \cos^2 \theta + B \cos \theta \sin \theta + C \sin^2 \theta, \quad (4-43a)$$

$$B''' = 2(C - A) \cos \theta \sin \theta + B(\cos^2 \theta - \sin^2 \theta), \quad (4-43b)$$

$$C''' = A \sin^2 \theta - B \cos \theta \sin \theta + C \cos^2 \theta, \quad (4-43c)$$

$$D''' = \xi \cos \theta + \eta \sin \theta, \quad (4-43d)$$

$$E''' = \eta \cos \theta - \xi \sin \theta, \quad (4-43e)$$

$$F''' = (m\xi + \eta n + \zeta)/2 \quad (4-43f)$$

и

$$\xi = 2Am + Bn + D, \quad (4-43g)$$

$$\eta = Bm + 2Cn + E, \quad (4-43h)$$

$$\zeta = Dm + En + 2F. \quad (4-43i)$$

Заметим, что это диагональная матрица, т.е. все недиагональные элементы равны нулю. Поворот устранил член $(B''' = 0)$, а перенос — линейные члены $(D''' = E''' = 0)$.

Угол вращения задается уравнением (4-40). Как и раньше, коэффициенты переноса можно получить, приравняв D''' и E''' к нулю, т.е. $\xi = \eta = 0$. Имеем

$$2Am + Bn + D = 0,$$

$$Bm + 2Cn + E = 0.$$

Отсюда получаем решение

$$m = \frac{2CD - BE}{B^2 - 4AC}, \quad (4-44a)$$

$$n = \frac{2AE - BD}{B^2 - 4AC}. \quad (4-44b)$$

Вспомним, что для центрального конического сечения $B^2 - 4AC \neq 0$. Запишем уравнение (4-41), используя (4-42):

$$\alpha x^2 + \beta y^2 = \kappa, \quad (4-45)$$

что является стандартным видом конического сечения. Остается систематически исследовать результат для различных значений α и β .

Если κ больше нуля, то α и β положительны, и сечение — эллипс. Если они имеют разные знаки и ни один из них не равен нулю, то сечение — гипербола.

Если α и β отрицательны, решения не существует.

Оба α и β одновременно не могут быть равны нулю, так как в этом случае уравнение (4-45) не содержит членов второго порядка. Однако один из коэффициентов α или β может быть нулевым. Пусть $\beta = 0$ (если $\alpha = 0$, замена x на y дает $\beta = 0$), тогда уравнение (4-45) принимает вид

$$\alpha x^2 = -\kappa.$$

Решением являются две параллельные линии $x = \pm\sqrt{-\kappa/\alpha}$ при $-\kappa/\alpha > 0$. Если $-\kappa/\alpha < 0$, то решения нет.

Если $\kappa = 0$, есть две возможности: α и β имеют одинаковые или разные знаки. В обоих случаях решение вырожденное. Если знаки α и β одинаковы, уравнению (4-45) удовлетворяет только начало координат $x = y = 0$. Можно считать, что это предельный случай эллипса.

Если знаки α и β различны, уравнение (4-45) принимает вид

$$y^2 = -\frac{\alpha}{\beta}x^2$$

или

$$y = \pm\sqrt{-\frac{\alpha}{\beta}}x,$$

т.е. это пара прямых, пересекающихся в начале координат, — предельный случай гиперболы.

Наконец, если $\beta = 0$ (если $\alpha = 0$, замена x на y дает $\beta = 0$), то решение — ось y для всех значений.

Для нецентрального сечения, параболы, оба линейных члена устраниить нельзя, однако можно убрать один линейный и один квадратичный член для x или y . Применим оператор переноса к уравнению (4-39) — квадратному уравнению после поворота:

$$[S^+] = \frac{1}{2} \begin{bmatrix} 2A^+ & 0 & D^+ \\ 0 & 2C^+ & E^+ \\ D^+ & E^+ & 2F^+ \end{bmatrix}, \quad (4-46)$$

где

$$A^+ = A'', \quad (4-47a)$$

$$C^+ = C'', \quad (4-47b)$$

$$D^+ = 2A''m + D'', \quad (4-47c)$$

$$E^+ = 2C''n + E'', \quad (4-47d)$$

$$F^+ = (m/2)D^+ + (n/2)E^+ + (mD'' + nE'' + 2F''). \quad (4-47e)$$

Угол поворота задан уравнением (4-40). Здесь либо $A^+ = A''$, либо $C^+ = C''$ обратится в нуль. Один из линейных членов для x или y устраняется, если D^+ или E^+ приравнять к нулю. Пусть $D^+ = 0$, тогда

$$m = \frac{-D''}{2A''}. \quad (4-48a)$$

Если $E^+ = 0$, то

$$n = \frac{-E''}{2C''}. \quad (4-48b)$$

Заметим, что при $A'' = 0$ значение m не определено, следовательно, устраняются только линейные члены относительно y . Если $C'' = 0$, не определено n и удаляются только члены с x .

Предположим, что уничтожены линейный y -член ($E^+ = 0$) и квадратичный x -член ($A^+ = 0$) (если $\beta = 0$, замена x и y дает $\gamma = 0$). Тогда $[S^+]$ принимает вид

$$[S^+] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \beta & 0 \\ \gamma & 0 & -\kappa \end{bmatrix}.$$

Запишем уравнение конического сечения

$$\beta y^2 + 2\gamma x = \kappa. \quad (4-49)$$

Чтобы привести параболу к стандартной ориентации с вершиной κ в центре координат, перенесем ее по оси x на

$$y^2 = -\frac{2\gamma}{\beta}x' = ax'.$$

Все вырожденные формы сечений центральны, т. е. парабола — это единственное нецентральное сечение. Результаты собраны в табл. 4-8.

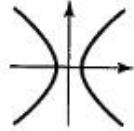
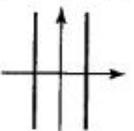
Пример 4-7 Сегмент гиперболы

Найти тип конического сечения, заданного формулой

$$2x^2 - 72xy + 23y^2 + 140x - 20y + 50 = 0,$$

изображенного непрерывной линией на рис. 4-16. Нарисовать его сегмент для $3 < x < 5$ для y_{min} . Сегмент должен быть нарисован с помощью параметрического представления из предыдущих разделов. Для того чтобы определить

Таблица 4-8 Конические сечения*

Название	Уравнение	Условия	Тип	Чертеж
Эллипс	$\alpha x^2 + \beta y^2 = \kappa$	$\kappa, \alpha, \beta > 0$	Центральный	
Гипербола	$\alpha x^2 + \beta y^2 = \kappa$	$\beta < 0 < \kappa, \alpha$	Центральный	
Парабола	$\alpha y^2 + \beta x = 0$ $\beta x^2 + \alpha y = 0$		Нецентральный	
Пустое множество	$\alpha x^2 + \beta y^2 = \kappa$	$\alpha, \beta < 0 < \kappa$	(Центральный)	
Точка	$\alpha x^2 + \beta y^2 = 0$	$\alpha, \beta > 0$	Центральный	
Пара прямых	$\alpha x^2 + \beta y^2 = 0$	$\beta < 0 < \alpha$	Центральный	
Параллельные прямые	$\alpha x^2 = \kappa$	$\alpha, \kappa > 0$	Центральный	
Пустое множество	$\alpha x^2 = \kappa$	$\alpha < 0 < \kappa$	(Центральный)	
«Повторяющаяся» прямая	$\alpha x^2 = 0$		Центральный	

* Эллипс, гипербола и парабола — это невырожденные сечения, остальные — вырожденные.

значения параметрического представления, используются методы, рассмотренные в данном разделе.

Сначала узнаем тип сечения

$$B^2 - 4AC = (-72)^2 - (4)(2)(23) = 5000 > 0,$$

т. е. это гипербола.

Приведем гиперболу к стандартному виду с помощью уравнений (4-40)–(4-44). Угол поворота

$$\theta = \frac{1}{2} \operatorname{arctg} \left(\frac{B}{A - C} \right) = \frac{1}{2} \operatorname{arctg} \left(\frac{-72}{2 - 23} \right) = \frac{1}{2} \operatorname{arctg} \left(\frac{24}{7} \right) = 36.87^\circ.$$

Подставим значения $\cos \theta = 4/5$, $\sin \theta = 3/5$ и получим

$$\begin{aligned} A''' &= A \cos^2 \theta + B \cos \theta \sin \theta + C \sin^2 \theta = \\ &= (2)(4/5)^2 + (-72)(4/5)(3/5) + (23)(3/5)^2 = \\ &= -625/25 = -25, \\ C''' &= A \sin^2 \theta - B \cos \theta \sin \theta + C \cos^2 \theta = \\ &= (2)(3/5)^2 - (-72)(4/5)(3/5) + (23)(4/5)^2 = \\ &= 1250/25 = 50. \end{aligned}$$

Коэффициенты переноса m и n таковы:

$$\begin{aligned} m &= \frac{2CD - BE}{B^2 - 4AC} = \frac{(2)(23)(140) - (-72)(-20)}{5000} = \frac{5000}{5000} = 1, \\ n &= \frac{2AE - BD}{B^2 - 4AC} = \frac{(2)(2)(-20) - (-72)(140)}{5000} = \frac{10000}{5000} = 2. \end{aligned}$$

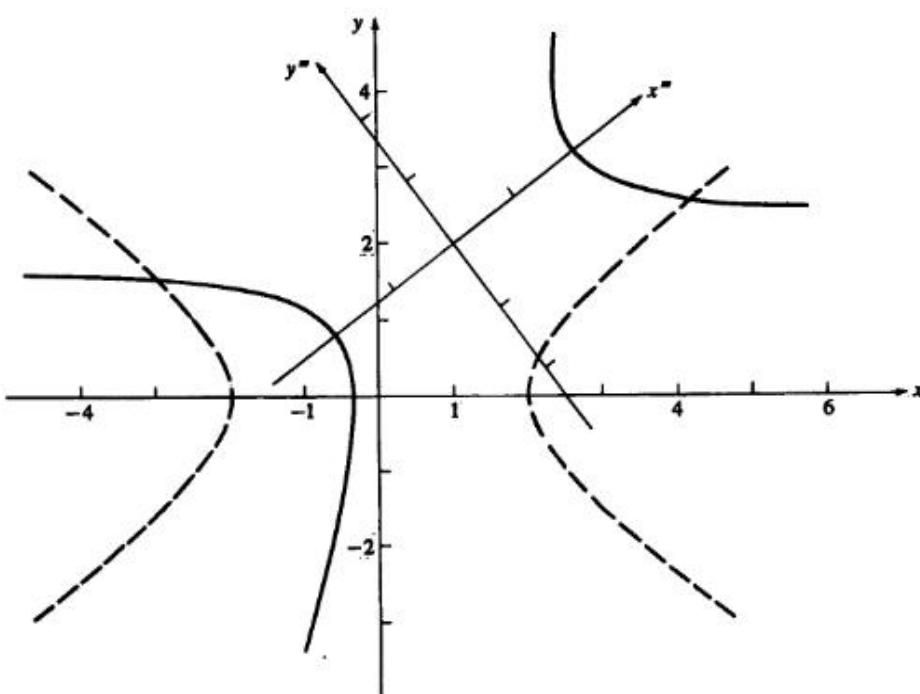


Рис. 4-16 Гипербола $2x^2 - 72xy + 23y^2 + 140x - 20y + 50 = 0$. Сплошная линия — заданная ориентация. Пунктир — стандартная ориентация.

Так как $\xi = \eta = 0$, константа равна

$$\begin{aligned} F''' &= \frac{\zeta}{2} = \frac{(Dm + En + 2F)}{2} \\ &= \frac{(140)(1) + (-20)(2) + (2)(50)}{2} = \frac{200}{2} = 100. \end{aligned}$$

Уравнение гиперболы

$$-25x^2 + 50y^2 + 100 = 0.$$

В стандартной форме $\alpha x^2 + \beta y^2 + \kappa = 0$ имеем

$$-x^2 + 2y^2 + 4 = 0 \Rightarrow x^2 - 2y^2 = 4.$$

В форме

$$\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1$$

имеем

$$\frac{x^2}{4} - \frac{y^2}{2} = 1,$$

что дает $a = 2$, $b = \sqrt{2}$.

Параметрическое представление параболы (4-16):

$$x = a \operatorname{ch}(t), \quad y = b \operatorname{sh}(t),$$

где t — параметр. Для параметрического изображения гиперболы необходимо найти величину t для $x = 3,5$. Заметим, что преобразования к стандартному виду переносят оси координат, а не само сечение. Поэтому соответствующие преобразованные значения x получаются обратными преобразованиями. В результате

$$\begin{aligned} [TR] &= [T][R] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -m & -n & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin(-\theta) & 0 \\ -\sin(-\theta) & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ -(m \cos \theta + n \sin \theta) & m \sin \theta - n \cos \theta & 1 \end{bmatrix} = \\ &= \begin{bmatrix} 4/5 & -3/5 & 0 \\ 3/5 & 4/5 & 0 \\ -2 & -1 & 1 \end{bmatrix}. \end{aligned}$$

Преобразованные координаты x

$$\begin{bmatrix} 3 & 2.871 & 1 \\ 5 & 2.476 & 1 \end{bmatrix} \begin{bmatrix} 4/5 & -3/5 & 0 \\ 3/5 & 4/5 & 0 \\ -2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 2.123 & -0.503 & 1 \\ 3.486 & -2.019 & 1 \end{bmatrix},$$

где также включена y координата. Тогда параметрические значения

$$t_{\min} = t|_{x=3} = \operatorname{Arch}\left(\frac{x}{2}\right) = \operatorname{Arch}\left(\frac{2.123}{2}\right) = 0.3486,$$

$$t_{\max} = t|_{x=5} = \operatorname{Arch}\left(\frac{x}{2}\right) = \operatorname{Arch}\left(\frac{3.486}{2}\right) = 1.1539.$$

Пользуясь этими значениями для t_{\min} и t_{\max} , получаем результат на рис. 4-16 и в табл. 4-9.

Таблица 4-9 Сегмент гиперболы в стандартной ориентации

t	x	y
0.349	2.123	-0.503
0.438	2.195	-0.640
0.528	2.285	-0.781
0.617	2.393	-0.929
0.706	2.520	-1.084
0.796	2.668	-1.248
0.885	2.837	-1.422
0.975	3.028	-1.608
1.064	3.244	-1.806
1.153	3.486	-2.019

Затем эти результаты обратным преобразованием переводятся в исходное положение:

$$[TR]^{-1} = \begin{bmatrix} 4/5 & 3/5 & 0 \\ -3/5 & 4/5 & 0 \\ 1 & 2 & 1 \end{bmatrix}.$$

Результат приведен в табл. 4-10 и на рис. 4-16.

Таблица 4-10 Сегмент гиперболы в исходной ориентации

t	x	y
0.349	3.000	2.871
0.438	3.140	2.805
0.528	3.297	2.746
0.617	3.472	2.693
0.706	3.667	2.645
0.796	3.883	2.602
0.885	4.123	2.564
0.975	4.387	2.531
1.064	4.679	2.502
1.153	5.000	2.476

4-11 ЛИТЕРАТУРА

- 4-1 Lancaster P. and Salkauskas K., *Curve and Surface Fitting, An Introduction*, Academic Press, 1987.
- 4-2 Smith L. B., «Drawing Ellipses, Hyperbolas, or Parabolas With a Fixed Number of Points and Maximum Inscribed Area», *Comp. J.*, Vol. 14, pp. 81–86, 1969.
- 4-3 Rogers D. F., «Interactive Graphics and Numerical Control», *Comp. Aid. Des.*, Vol. 12, pp. 253–261, 1980.
- 4-4 Middleditch A. E. and Stacey T. W., «Robust Computation of Tangent Lines and Circles», *Comp. Aid. Des.*, Vol. 19, pp. 503–507, 1987.

ПРОСТРАНСТВЕННЫЕ КРИВЫЕ

5-1 ВВЕДЕНИЕ

Трехмерные, или пространственные, кривые широко используются в проектировании и разработке самой различной продукции: автомобилей, кораблей, самолетов, обуви, бутылок, зданий и т. д. Также они имеют большое значение для описания и интерпретации физических явлений в геологии, физике и медицине.

До начала применения математических и компьютерных моделей в процессе производства, дизайна и изготовления использовалась начертательная геометрия. Многие ее методы были перенесены в машинную графику.

Поверхности часто изображаются как сеть кривых, лежащих в ортогональных секущих плоскостях, с трехмерными контурами деталей. Пример приведен на рис. 5-1. Здесь сечения были получены оцифровкой физической модели или

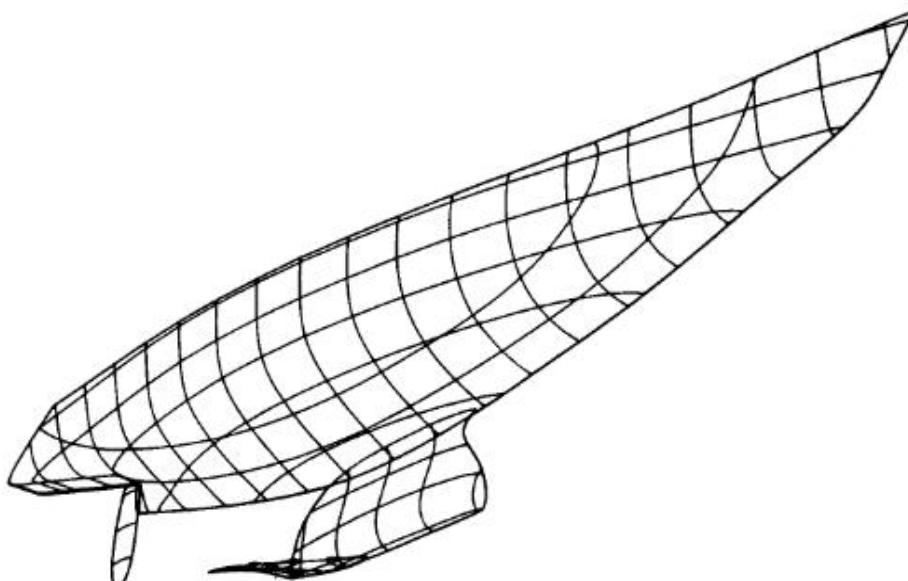


Рис. 5-1 Поверхность в виде сети ортогональных плоских кривых (Дж. Хейзен)

чертежа и математическим подбором кривой, проходящей через все заданные точки. В данной главе рассматриваются два таких метода: кубические сплайны и параболическая интерполяция. Существуют и другие реализации этого подхода, рассмотренные, например, в работе [5-1].

Другой подход состоит в том, что математическое описание кривых генерируется без изначального знания формы кривой. Его примеры — это кривые Безье и их обобщение до B-сплайнов. Эти методы отличаются тем, что кривая может не проходить ни через одну заданную точку. Контрольные точки определяют только направление изгиба.

Как будет показано далее, оба подхода можно применить при любом способе задания кривой.

5-2 ПРЕДСТАВЛЕНИЕ ПРОСТРАНСТВЕННЫХ КРИВЫХ

Трехмерные кривые можно представить параметрически или непараметрически. Явное непараметрическое представление имеет вид

$$\begin{aligned}x &= x, \\y &= f(x), \\z &= g(x).\end{aligned}$$

Неявное непараметрическое представление кривой как пересечения двух поверхностей задается уравнениями:

$$\begin{aligned}f(x, y, z) &= 0, \\g(x, y, z) &= 0.\end{aligned}$$

Пример 5-1 Пространственная кривая

Найти линию пересечения двух поверхностей второго порядка

$$\begin{aligned}f(x, y, z) &= y - z^2 = 0, \\g(x, y, z) &= zx - y^2 = 0.\end{aligned}$$

При условиях что

$$\det \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix} \neq 0,$$

$z \neq 0$, x и y можно выразить относительно z и получить явный вид линии пересечения

$$\begin{aligned}y &= z^2, \\x &= \frac{y^2}{z} = z^3.\end{aligned}$$

Заметим, что при пересечении двух поверхностей второго порядка получается кривая третьего порядка. Поверхности и кривая пересечения показаны на рис. 5-2.

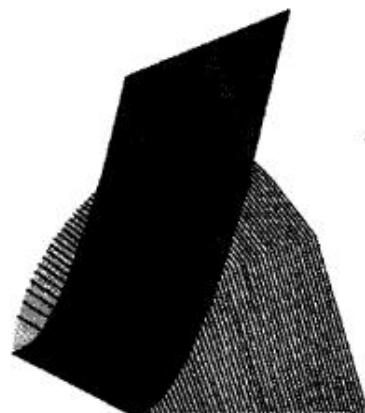


Рис. 5-2 Пересечение кривых второго порядка.

Общий параметрический вид пространственной кривой можно записать в виде:

$$\begin{aligned}x &= x(t), \\y &= y(t), \\z &= z(t),\end{aligned}$$

где параметр t изменяется в определенных пределах $t_1 \leq t \leq t_2$. В приведенном выше явном непараметрическом представлении x можно рассматривать как параметр, $x = t$. Тогда эта же кривая имеет параметрическую форму

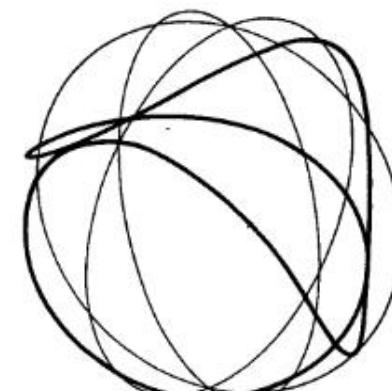
$$\begin{aligned}x &= t, \\y &= f(t), \\z &= g(t).\end{aligned}$$

Далее, пусть $z = t$ в неявном непараметрическом представлении из примера 5-1, тогда

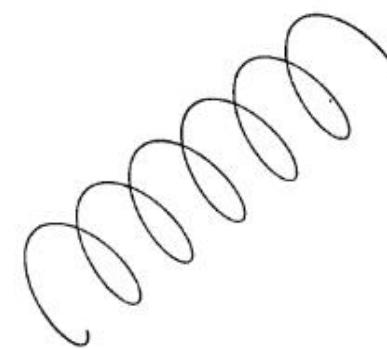
$$\begin{aligned}x &= t^3, \\y &= t^2, \\z &= t.\end{aligned}$$

Некоторые полезные параметрические трехмерные кривые имеют известное аналитическое решение. Например, кривая шва на теннисном или бейсбольном мяче имеет вид:

$$\begin{aligned}x &= \lambda[a \cos(\theta + \frac{\pi}{4}) - b \cos 3(\theta + \frac{\pi}{4})], \\y &= \mu[a \sin(\theta + \frac{\pi}{4}) - b \sin 3(\theta + \frac{\pi}{4})], \\z &= c \sin(2\theta),\end{aligned}$$



(a)



(b)

Рис. 5-3 Примеры параметрических пространственных кривых. (a) Шов бейсбольного мяча; (b) круговая спираль.

где

$$\begin{aligned}\lambda &= 1 + d \sin(2\theta) = 1 + d \left(\frac{z}{c}\right), \\ \mu &= 1 - d \sin(2\theta) = 1 - d \left(\frac{z}{c}\right),\end{aligned}$$

параметр $\theta = 2\pi t$ и $0 \leq t \leq 1.0$. Если $d = 0$ и $c^2 = 4ab$, то кривая лежит на сфере радиуса $a + b$. На рис. 5-3а приведен пример для $a = 1$, $b = 1$, $c = 2$, $d = 0$, где кривая лежит на сфере радиуса 2.

Другой пример параметрической пространственной кривой — круговая спираль:

$$\begin{aligned}x &= r \cos t, \\y &= r \sin t, \\z &= bt\end{aligned}$$



Рис. 5-4 Физические отвесы и сплайн.

для $r, b \neq 0, -\infty < t < \infty$. Эта кривая лежит на поверхности цилиндра радиуса $|r|$. Уравнение $z = bt$ отвечает за распространение спирали по оси z . После каждого изменения параметра t на 2π переменные x и y возвращаются к своим первоначальным значениям, а z увеличивается или уменьшается на $2\pi|b|$ в зависимости от знака b . Эта величина называется шагом спирали. Пример изображен на рис. 5-3б.

5-3 КУБИЧЕСКИЕ СПЛАЙНЫ

В промышленном производстве, например, судо-, автомобилье- и авиастроении, окончательная форма в реальном или близком к нему масштабе определяется в процессе доводки¹.

Автоматизация этого процесса представляла значительный интерес для машинной графики. Форма математического сплайна повторяет контур физического сплайна (рис. 5-4), т. е. гибкой деревянной или пластмассовой линейки, проходящей через определенные точки. Для изменения формы сплайна используются свинцовые грузики. Меняя их количество и расположение, получившуюся кривую стараются сделать более гладкой, красивой и «приятной для глаза»².

Если рассматривать физический сплайн как тонкую гибкую рейку, его форма (отклонение y) определяется уравнением Эйлера (5-2) для момента изгиба $M(x)$ вдоль рейки:

$$M(x) = \frac{EI}{R(x)},$$

где E — модуль Юнга, зависящий от свойств материала рейки, I — момент инерции, определяемый формой кривой, $R(x)$ — радиус кривизны.

¹ В судостроении, например, необходимо определить форму судна более точно, чем это возможно на небольшом чертеже. Поэтому общепринято изображать контур судна в натуральную величину на плазе.

² Не существует общепринятого термина для оценки внешнего вида кривой. Мы не пытаемся дать математического определения, потому что любая такая попытка вызовет волну протеста со стороны специалистов.

Для малых отклонений ($y' \ll 1$) радиус приближенно равен

$$\frac{1}{R(x)} = \frac{y''}{(1 + y'^2)^{3/2}} \approx y'',$$

где штрих обозначает производную по x — расстоянию вдоль рейки, а y — отклонение рейки. Уравнение Эйлера принимает вид

$$y'' = \frac{M(x)}{EI}.$$

Пусть грузики действуют как простые подпорки, тогда момент изгиба между ними изменяется линейно. Подставляя $M(x) = Ax + B$ в уравнение Эйлера, получаем

$$y'' = \frac{Ax + B}{EI}$$

и после двойного интегрирования

$$y = A_1x^3 + B_1x^2 + C_1x + D_1.$$

Таким образом, форма сплайна задается кубическим полиномом.

В общем случае математический сплайн это кусочный полином степени K с непрерывной производной степени $K-1$ в точках соединения сегментов. Так, например, кубический сплайн имеет в точках соединения непрерывность второго порядка. Кусочные сплайны из многочленов невысокого порядка очень удобны для интерполяции кривых, так как они не требуют больших вычислительных затрат и не вызывают численных отклонений, свойственных многочленам высокого порядка. По аналогии с физическими сплайнами обычно используется серия кубических сегментов, причем каждый сегмент проходит через две точки. Кубический сплайн удобен еще и тем, что это кривая наименьшего порядка, допускающая точки перегиба и изгиб в пространстве.

Уравнение одного параметрического сегмента сплайна таково:

$$P(t) = \sum_{i=1}^4 B_i t^{i-1}, \quad t_1 \leq t \leq t_2, \quad (5-1)$$

где t_1 и t_2 — значения параметров в начале и конце сегмента. $P(t)$ — вектор к любой точке сегмента. $P(t) = [x(t) \ y(t) \ z(t)]$ — это векторно-значная функция, где три составляющие $P(t)$ — декартовы координаты вектора¹.

¹ $P(t) = [r(t) \ \theta(t) \ z(t)]$, где $r(t)$, $\theta(t)$, $z(t)$ — компоненты цилиндрической системы координат, также совершенно справедливо, как, впрочем, и представление в любой другой системе координат. (См. задачу 5-8.)



Рис. 5-5 Один сегмент кубического сплайна.

Каждая составляющая имеет вид, похожий на $P(t)$, т. е.

$$x(t) = \sum_{i=1}^4 B_{i_x} t^{i-1}, \quad t_1 \leq t \leq t_2,$$

$$y(t) = \sum_{i=1}^4 B_{i_y} t^{i-1}, \quad t_1 \leq t \leq t_2,$$

$$z(t) = \sum_{i=1}^4 B_{i_z} t^{i-1}, \quad t_1 \leq t \leq t_2.$$

Постоянные коэффициенты B_i вычисляются исходя из четырех граничных условий для сегмента сплайна. Запишем уравнение (5-1) в виде

$$P(t) = B_1 + B_2 t + B_3 t^2 + B_4 t^3, \quad t_1 \leq t \leq t_2. \quad (5-2)$$

Пусть P_1 и P_2 — векторы концов сегмента (см. рис. 5-5). Пусть также P'_1 и P'_2 , производные по t , будут касательными векторами в концах сегмента. Дифференцируя уравнение (5-1), получим

$$P'(t) = [x'(t) \quad y'(t) \quad z'(t)] = \sum_{i=1}^4 B_i (i-1) t^{i-2}, \quad t_1 \leq t \leq t_2. \quad (5-3)$$

Запишем результат

$$P'(t) = B_2 + 2B_3 t + 3B_4 t^2, \quad t_1 \leq t \leq t_2. \quad (5-4)$$

Предположим, без потери общности, что $t_1 = 0$, и применим граничные условия

$$P(0) = P_1, \quad (5-5a)$$

$$P(t_2) = P_2, \quad (5-5b)$$

$$P'(0) = P'_1, \quad (5-5c)$$

$$P'(t_2) = P'_2. \quad (5-5d)$$

Получим четыре уравнения для неизвестных B_i :

$$P(0) = B_1 = P_1, \quad (5-6a)$$

$$P'(0) = \sum_{i=1}^4 (i-1) t^{i-2} B_i \Big|_{t=0} = B_2 = P'_1, \quad (5-6b)$$

$$P(t_2) = \sum_{i=1}^4 B_i t^{i-1} \Big|_{t=t_2} = B_1 + B_2 t_2 + B_3 t_2^2 + B_4 t_2^3, \quad (5-6c)$$

$$P'(t_2) = \sum_{i=1}^4 (i-1) t^{i-2} B_i \Big|_{t=t_2} = B_2 + 2B_3 t_2 + 3B_4 t_2^2. \quad (5-6d)$$

Решения для B_3 и B_4 имеют вид:

$$B_3 = \frac{3(P_2 - P_1)}{t_2^2} - \frac{2P'_1}{t_2} - \frac{P'_2}{t_2} \quad (5-7a)$$

$$\text{и} \quad B_4 = \frac{2(P_1 - P_2)}{t_2^3} + \frac{P'_1}{t_2^2} + \frac{P'_2}{t_2^2}. \quad (5-7b)$$

Величины B_1 , B_2 , B_3 и B_4 задают сегмент кубического сплайна. Очевидно, что форма сегмента зависит от положения и касательных векторов в концах сегмента. Далее, заметим, что в результатах присутствует значение параметра $t = t_2$ в конце сегмента. Так как каждая конечная точка и вектор касания имеют три компоненты, параметрическое уравнение кубической пространственной кривой зависит от двенадцати векторных компонент и значения параметра t_2 в конце сегмента.

Подставив уравнения (5-6) и (5-7) в (5-1), получим уравнение для одного сегмента кубического сплайна:

$$P(t) = P_1 + P'_1 t + \left[\frac{3(P_2 - P_1)}{t_2^2} - \frac{2P'_1}{t_2} - \frac{P'_2}{t_2} \right] t^2 + \left[\frac{2(P_1 - P_2)}{t_2^3} + \frac{P'_1}{t_2^2} + \frac{P'_2}{t_2^2} \right] t^3. \quad (5-8)$$

Это уравнение для одного сегмента. Чтобы получить кривую целиком, нужно соединить множество сегментов. На рис. 5-6 показаны два соседних сегмента. Если известны векторы P_1 , P_2 , P_3 , касательные векторы P'_1 , P'_2 , P'_3 и значения параметров t_2 , t_3 , то форма каждого сегмента определяется из уравнения (5-8). Однако маловероятно, что известен касательный вектор P'_2 в точке соединения. К счастью, его можно вывести из условия непрерывности.

Вспомним, что кусочный сплайн степени K имеет непрерывность степени $K-1$ в точках соединения; непрерывность кубического сплайна равна двум. Для этого должна быть непрерывна вторая производная $P''_2(t)$ или кривизна линии. Дважды проинтегрировав уравнение (5-1), получим

$$P''(t) = \sum_{i=1}^4 (i-1)(i-2) B_i t^{i-3}, \quad t_1 \leq t \leq t_2. \quad (5-9)$$

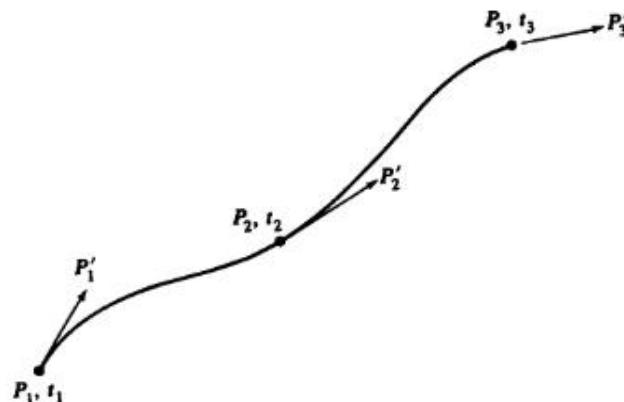


Рис. 5-6 Два кусочно кубических сегмента сплайна.

Для первого куска сплайна параметр изменяется в пределах $0 \leq t \leq t_2$. Подставим $t = t_2$ в уравнение (5-9):

$$P'' = 6B_4t_2 + 2B_3.$$

Для второго участка сплайна параметр изменяется в диапазоне $0 \leq t \leq t_3$. Подставим в уравнение (5-9) значение $t = 0$ в начале второго участка

$$P'' = 2B_3.$$

Приравнивая полученные результаты и пользуясь уравнениями (5-6a, b) и (5-7a), получим

$$\begin{aligned} 6t_2 \left[\frac{2(P_1 - P_2)}{t_2^3} + \frac{P'_1}{t_2^2} + \frac{P'_2}{t_2^2} \right] + 2 \left[\frac{3(P_2 - P_1)}{t_2^2} - \frac{2P'_1}{t_2} - \frac{P'_2}{t_2} \right] = \\ = 2 \left[\frac{3(P_3 - P_2)}{t_3^2} - \frac{2P'_2}{t_3} - \frac{P'_3}{t_3} \right]. \end{aligned}$$

Левая часть этого уравнения представляет кривизну в конце первого сегмента, а правая — в начале второго. Домножим на $t_2 t_3$ и сгруппируем члены:

$$t_3 P'_1 + 2(t_3 + t_2) P'_2 + t_2 P'_3 = \frac{3}{t_2 t_3} [t_2^2 (P_3 - P_2) + t_3^2 (P_2 - P_1)]. \quad (5-10)$$

Отсюда определяется P'_2 , неизвестный касательный вектор в точке соединения. Отметим, что в окончательном уравнении опять присутствуют значения параметра t в концах сегментов t_2 и t_3 .

Полученную формулу можно обобщить для n точек, и для $n - 1$ сегментов кубического сплайна получить непрерывность второго порядка в точках соедине-

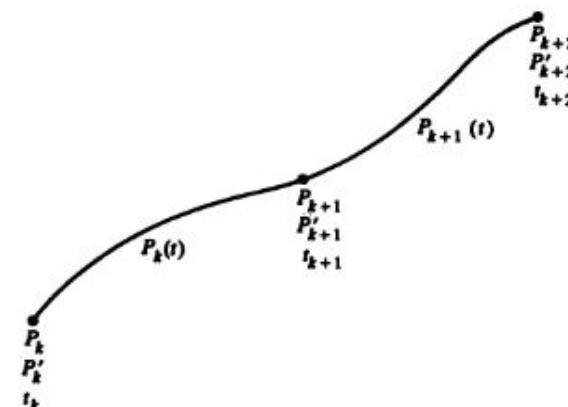


Рис. 5-7 Обозначения множества кусочно кубических сегментов сплайна.

ния. Обобщенное уравнение для двух любых соседних сегментов сплайна $P_k(t)$ и $P_{k+1}(t)$ ¹ в обозначениях рис. 5-7 имеет вид:

$$\begin{aligned} P_k(t) = P_k + P'_k t + \left[\frac{3(P_{k+1} - P_k)}{t_{k+1}^2} - \frac{2P'_k}{t_{k+1}} - \frac{P'_{k+1}}{t_{k+1}} \right] t^2 + \\ + \left[\frac{2(P_k - P_{k+1})}{t_{k+1}^3} + \frac{P'_k}{t_{k+1}^2} + \frac{P'_{k+1}}{t_{k+1}^2} \right] t^3 \end{aligned} \quad (5-11)$$

для первого сегмента и

$$\begin{aligned} P_{k+1}(t) = P_{k+1} + P'_{k+1} t + \left[\frac{3(P_{k+2} - P_{k+1})}{t_{k+2}^2} - \frac{2P'_{k+1}}{t_{k+2}} - \frac{P'_{k+2}}{t_{k+2}} \right] t^2 + \\ + \left[\frac{2(P_{k+1} - P_{k+2})}{t_{k+2}^3} + \frac{P'_{k+1}}{t_{k+2}^2} + \frac{P'_{k+2}}{t_{k+2}^2} \right] t^3 \end{aligned} \quad (5-12)$$

для второго, так как для каждого сегмента параметр начинает изменяться с нуля, для первого $0 \leq t \leq t_{k+1}$ и для второго $-0 \leq t \leq t_{k+2}$.

Приравнивание вторых производных в точках стыковки для любых соседних сегментов, $P''_k(t_k) = P''_{k+1}(0)$, дает общий результат, эквивалентный уравнению (5-10),

$$\begin{aligned} t_{k+2} P'_k + 2(t_{k+1} + t_{k+2}) P'_{k+1} + t_{k+1} P'_{k+2} = \\ = \frac{3}{t_{k+1} t_{k+2}} [t_{k+1}^2 (P_{k+2} - P_{k+1}) + t_{k+2}^2 (P_{k+1} - P_k)], \quad 1 \leq k \leq n - 2, \end{aligned} \quad (5-13)$$

откуда определяется касательный вектор в точках соединения любых двух сегментов P_k и P_{k+1} .

¹ Чтобы не вводить лишних определений, условимся обозначать участок параметрической кривой между точками P_k и P_{k+1} как $P_k(t)$.

Рекурсивное использование уравнения (5-13) для всех сегментов сплайна порождает $n - 2$ уравнений касательных векторов P'_k , $2 \leq k \leq n - 1$. В матричной форме:

$$\begin{bmatrix} t_3 & 2(t_2 + t_3) & t_2 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & t_4 & 2(t_3 + t_4) & t_3 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & t_5 & 2(t_4 + t_5) & t_4 & 0 & \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & t_n & 2(t_n + t_{n-1}) & t_{n-1} & \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ \vdots \\ P'_n \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{3}{t_2 t_3} \{t_2^2 (P_3 - P_2) + t_3^2 (P_2 - P_1)\} \\ \frac{3}{t_3 t_4} \{t_3^2 (P_4 - P_3) + t_4^2 (P_3 - P_2)\} \\ \vdots \\ \frac{3}{t_{n-1} t_n} \{t_{n-1}^2 (P_n - P_{n-1}) + t_n^2 (P_{n-1} - P_{n-2})\} \end{bmatrix} \quad (5-14)$$

$$= \begin{bmatrix} P'_1 \\ \frac{3}{t_2 t_3} \{t_2^2 (P_3 - P_2) + t_3^2 (P_2 - P_1)\} \\ \frac{3}{t_3 t_4} \{t_3^2 (P_4 - P_3) + t_4^2 (P_3 - P_2)\} \\ \vdots \\ \frac{3}{t_{n-1} t_n} \{t_{n-1}^2 (P_n - P_{n-1}) + t_n^2 (P_{n-1} - P_{n-2})\} \\ P'_n \end{bmatrix} \quad (5-15)$$

или

$$[M][P'] = [R],$$

где матрица $[M]$ квадратная и обратимая. Заметим также, что $[M]$ трехдиагональная¹, что снижает вычислительные затраты на ее обращение. Далее, матрица $[M]$ диагонально доминантная². Отсюда следует, что у нее существует единственное решение:

$$[P'] = [M]^{-1}[R]. \quad (5-16)$$

Если нам известны P'_k , то легко определить коэффициенты B_i для каждого сегмента сплайна. Обобщая уравнения (5-6)–(5-11), получим

$$\begin{aligned} B_{1k} &= P_k, \\ B_{2k} &= P'_k, \\ B_{3k} &= \frac{3(P_{k+1} - P_k)}{t_{k+1}^2} - \frac{2P'_k}{t_{k+1}} - \frac{P'_{k+1}}{t_{k+1}}, \\ B_{4k} &= \frac{2(P_k - P_{k+1})}{t_{k+1}^3} + \frac{P'_k}{t_{k+1}^2} + \frac{P'_{k+1}}{t_{k+1}^2}. \end{aligned}$$

Так как P_k и P'_k – это векторные величины, то и B_i тоже векторные; если P_k и P'_k имеют x, y, z составляющие, значит, и B_i также имеет эти составляющие.

¹ У трехдиагональной матрицы элементы расположены только на главной диагонали, первой верхней и первой нижней диагоналях.

² У диагонально доминантной матрицы порядок членов на главной диагонали больше, чем других членов в этом же ряду.

$$[M^*][P'] = [R].$$

Матрица $[M^*]$ неквадратная, так как имеется только $n - 2$ уравнений для n векторов, и ее нельзя обратить и получить решение для $[P']$. Если предположить, что известны касательные векторы на концах кривой P'_1 и P'_n , проблема разрешается. Теперь матрица имеет вид

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_3 & 2(t_2 + t_3) & t_2 & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & t_4 & 2(t_3 + t_4) & t_3 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & t_5 & 2(t_4 + t_5) & t_4 & 0 & \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & t_n & 2(t_n + t_{n-1}) & t_{n-1} & \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ \vdots \\ P'_n \\ 0 \\ 1 \end{bmatrix} =$$

В матричной форме уравнение любого сегмента сплайна k таково:

$$\begin{aligned} [B] &= \begin{bmatrix} B_{1k} \\ B_{2k} \\ B_{3k} \\ B_{4k} \end{bmatrix} = \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{3}{t_{k+1}^2} & -\frac{2}{t_{k+1}} & \frac{3}{t_{k+1}^2} & -\frac{1}{t_{k+1}} \\ \frac{t_{k+1}^3}{t_{k+1}^2} & \frac{t_{k+1}^2}{t_{k+1}} & -\frac{t_{k+1}^3}{t_{k+1}^2} & \frac{1}{t_{k+1}} \end{bmatrix} \begin{bmatrix} P_k \\ P'_k \\ P_{k+1} \\ P'_{k+1} \end{bmatrix}. \end{aligned} \quad (5-17)$$

Пусть требуется задать кубический сплайн, проходящий через n точек P_k , $1 \leq k \leq n$ с касательными векторами на концах P'_1 и P'_n . Из уравнения (5-16) находим внутренние касательные векторы P'_k , $2 \leq k \leq n-1$. Затем из уравнения (5-17) с известными координатами концов каждого сегмента и касательными векторами определяются B_{ik} , $1 \leq i \leq 4$, для каждого сегмента. Окончательное обобщение уравнения (5-1)

$$P_k(t) = \sum_{i=1}^4 B_{ik} t^{i-1}, \quad 0 \leq t \leq t_{k+1}, \quad 1 \leq k \leq n-1, \quad (5-18)$$

используется для расчета сегмента сплайна.

В матричном виде уравнение (5-18) выглядит следующим образом:

$$P_k(t) = [1 \quad t \quad t^2 \quad t^3] \begin{bmatrix} B_{1k} \\ B_{2k} \\ B_{3k} \\ B_{4k} \end{bmatrix}, \quad 0 \leq t \leq t_{k+1}. \quad (5-19)$$

Подставляя уравнение (5-17) и перегруппируя члены, получим

$$P_k(\tau) = [F_1(\tau) \quad F_2(\tau) \quad F_3(\tau) \quad F_4(\tau)] \begin{bmatrix} P_k \\ P_{k+1} \\ P'_k \\ P'_{k+1} \end{bmatrix}, \quad \begin{array}{l} 0 \leq \tau \leq 1 \\ 1 \leq k \leq n-1 \end{array}, \quad (5-20)$$

где

$$\tau = (t/t_{k+1}),$$

$$F_{1k}(\tau) = 2\tau^3 - 3\tau^2 + 1, \quad (5-21a)$$

$$F_{2k}(\tau) = -2\tau^3 + 3\tau^2, \quad (5-21b)$$

$$F_{3k}(\tau) = \tau(\tau^2 - 2\tau + 1)t_{k+1}, \quad (5-21c)$$

$$F_{4k}(\tau) = \tau(\tau^2 - \tau)t_{k+1} \quad (5-21d)$$

называются весовыми функциями.

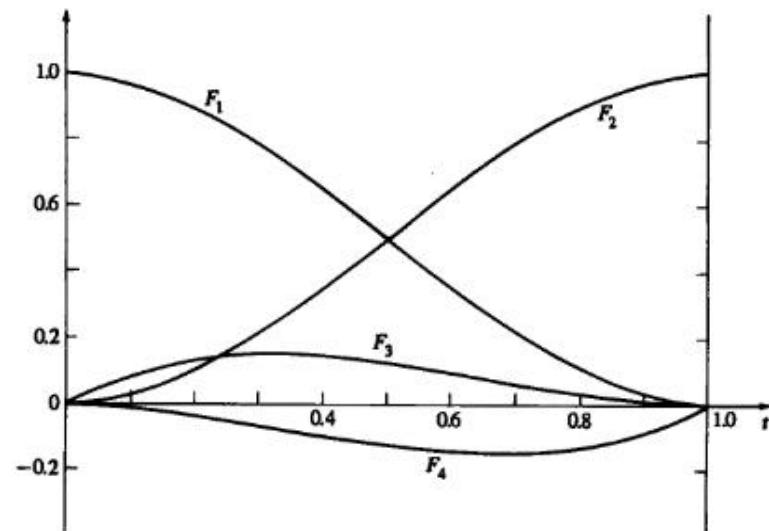


Рис. 5-8 Весовые функции кубического сплайна для $t_{k+1} = 1.0$

Пользуясь этими определениями, запишем уравнение (5-20) в матричном виде

$$P_k(\tau) = [F][G], \quad (5-22)$$

где $[F]$ — матрица весовой функции

$$[F] = [F_1(\tau) \quad F_2(\tau) \quad F_3(\tau) \quad F_4(\tau)] \quad (5-23)$$

и

$$[G]^T = [P_k \quad P_{k+1} \quad P'_k \quad P'_{k+1}] \quad (5-24)$$

содержит геометрическую информацию. Как будет видно из дальнейшего, уравнения типа (5-22), т. е. матрица весовой функции, умноженная на матрицу геометрических условий, часто применяются для описания кривых и поверхностей.

Из уравнения (5-21) видно, что каждая весовая функция имеет третий порядок. Любая точка на сегменте кубического сплайна это взвешенная сумма конечных точек и касательных векторов. Коэффициенты F_{ik} выступают в роли весовых функций. На рис. 5-8 изображены F_i для $t_{k+1} = 1.0$. Из рисунка видно, что $F_1(0) = 1$ и $F_2(0) = F_3(0) = F_4 = 0$, т. е. кривая проходит через вектор-точку P_1 . Аналогично $F_2(1) = 1$ и $F_1(1) = F_3(1) = F_4(1) = 0$, т. е. кривая также проходит через вектор-точку P_2 . Далее отметим симметрию F_1 и F_2 , и F_3 и F_4 . Фактически $F_2(\tau) = 1 - F_1(\tau)$. Наконец, обратим внимание на относительный порядок F_1 , F_2 , F_3 и F_4 . Значительная разница величин говорит о том, что в общем случае положение конечных точек имеет большее влияние, чем касательные векторы.

Вспомним, что кусочный кубический сплайн определяется точками, векторами касательных и величинами параметра, т. е. t_k в концах всех сегментов. Выбор t_k влияет на гладкость кривой.

Непрерывность второй производной в точках внутреннего соединения сама по себе не обеспечивает гладкости кривой в смысле минимальности кривизны вдоль нее. Подбирая соответствующие значения, можно минимизировать коэффициенты B_3 и B_4 для каждого сегмента и достичь большей гладкости кривой. Обычно эти дополнительные вычисления не требуются. Для практических целей применяются более простые методы, наподобие рассмотренных здесь.

Один метод вычисления t_k — установить величины параметров равными длиной хорд между соседними точками. При этом качество кривой удовлетворяет требованиям большинства прикладных задач. Другой метод состоит в том, что для нормализации вариации t_k полагается равным единице у каждого сегмента сплайна. Такой выбор упрощает вычисления (см. разд. 5-4). Как видно из приведенных выше уравнений, любой выбор t_k приводит к другим коэффициентам, и, следовательно, получаются различные кривые, проходящие через заданные точки.

Рассмотрим пример¹.

Пример 5-2 Кубический сплайн

Пусть даны четыре вектор-точки на плоскости: $P_1 [0 \ 0]$, $P_2 [1 \ 1]$, $P_3 [2 \ -1]$, $P_4 [3 \ 0]$ (см. рис. 5-9). Найти кусочный кубический сплайн, проходящий через них, используя хордовую аппроксимацию t_k . Касательные векторы в концах: $P'_1 [1 \ 1]$ и $P'_4 [1 \ 1]$. Найти промежуточные точки при $\tau = 1/3, 2/3$ для каждого сегмента.

Сначала найдем

$$t_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = \sqrt{(1)^2 + (1)^2} = \sqrt{2},$$

$$t_3 = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} = \sqrt{(1)^2 + (-2)^2} = \sqrt{5},$$

$$t_4 = \sqrt{(x_4 - x_3)^2 + (y_4 - y_3)^2} = \sqrt{(1)^2 + (1)^2} = \sqrt{2}.$$

Внутренние касательные векторы P'_2 и P'_3 вычисляются из уравнения (5-15):

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ t_3 & 2(t_2 + t_3) & t_2 & 0 \\ 0 & t_4 & 2(t_3 + t_4) & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \end{bmatrix} = \begin{bmatrix} P'_1 \\ \frac{3}{t_2 t_3} \{t_2^2 (P_3 - P_2) + t_3^2 (P_2 - P_1)\} \\ \frac{3}{t_3 t_4} \{t_3^2 (P_4 - P_3) + t_4^2 (P_3 - P_2)\} \\ P'_4 \end{bmatrix}.$$

¹ На протяжении всей этой главы рассматриваются двумерные примеры. Это сделано только для упрощения выкладок и представления результатов. Третье измерение вводится простым расширением (см. вопросы к гл. 5).

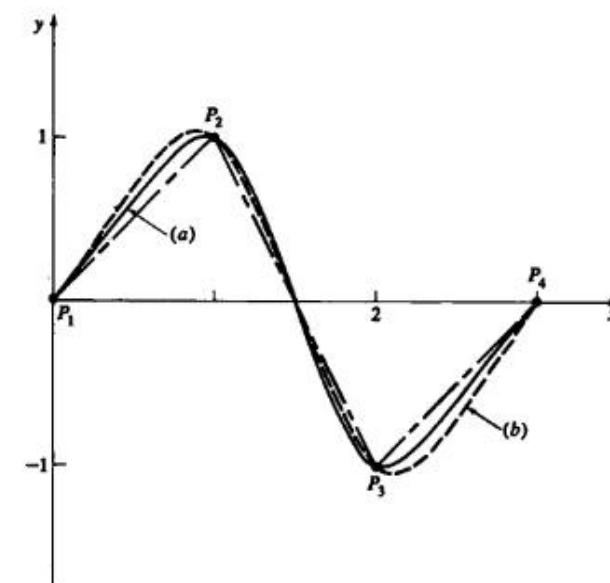


Рис. 5-9 Кусочный кубический сплайн. (a) t_k вычислены с помощью хордовой аппроксимации; (b) t_k нормализованы к 1.

Сделав подстановку, получим

$$\begin{aligned} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ \sqrt{5} & 2(\sqrt{2} + \sqrt{5}) & \sqrt{2} & 0 \\ 0 & \sqrt{2} & 2(\sqrt{5} + \sqrt{2}) & \sqrt{5} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \end{bmatrix} = \\ & = \begin{bmatrix} [1 \ 1] \\ \frac{3}{\sqrt{2}\sqrt{5}} \{2([2 \ -1] - [1 \ 1]) + 5([1 \ 1] - [0 \ 0])\} \\ \frac{3}{\sqrt{5}\sqrt{2}} \{5([3 \ 0] - [2 \ -1]) + 2([2 \ -1] - [1 \ 1])\} \\ [1 \ 1] \end{bmatrix} = \\ & = \begin{bmatrix} [1 \ 1] \\ \frac{3}{\sqrt{2}\sqrt{5}} \{[2 \ -4] + [5 \ 5]\} \\ \frac{3}{\sqrt{5}\sqrt{2}} \{[5 \ 5] + [2 \ -4]\} \\ [1 \ 1] \end{bmatrix} = \begin{bmatrix} 0.949 & [1 \ 1] \\ 0.949 & [7 \ 1] \\ 7 & 1 \end{bmatrix} \end{aligned}$$

или

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2.236 & 7.300 & 1.414 & 0 \\ 0 & 1.414 & 7.300 & 2.236 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 6.641 & 0.949 \\ 6.641 & 0.949 \\ 1 & 1 \end{bmatrix}.$$

С помощью инверсии и умножения вычисляются касательные векторы

$$\begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.318 & 0.142 & -0.028 & 0.062 \\ 0.062 & -0.028 & 0.142 & -0.318 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 6.641 & 0.949 \\ 6.641 & 0.949 \\ 1 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 \\ 0.505 & -0.148 \\ 0.505 & -0.148 \\ 1 & 1 \end{bmatrix}.$$

Из уравнения (5-21) находим весовые функции первого сегмента:

$$F_1(1/3) = 2(1/3)^3 - 3(1/3)^2 + 1 = \frac{20}{27} = 0.741,$$

$$F_2(1/3) = -2(1/3)^3 + 3(1/3)^2 = \frac{7}{27} = 0.259,$$

$$F_3(1/3) = (1/3)[(1/3)^2 - 2(1/3) + 1]\sqrt{2} = \frac{4\sqrt{2}}{27} = 0.210,$$

$$F_4(1/3) = (1/3)[(1/3)^2 - 1/3]\sqrt{2} = -\frac{2\sqrt{2}}{27} = -0.105$$

и второго:

$$F_1(2/3) = 2(2/3)^3 - 3(2/3)^2 + 1 = \frac{7}{27} = 0.259,$$

$$F_2(2/3) = -2(2/3)^3 + 3(2/3)^2 = \frac{20}{27} = 0.741,$$

$$F_3(2/3) = (2/3)[(2/3)^2 - 2(2/3) + 1]\sqrt{2} = \frac{2\sqrt{2}}{27} = 0.105,$$

$$F_4(2/3) = (2/3)[(2/3)^2 - 2/3]\sqrt{2} = -\frac{4\sqrt{2}}{27} = -0.210.$$

Точка на первом сегменте сплайна при $\tau = 1/3$ получается из уравнения (5-22), $P(\tau) = [F][G]$, т. е.

$$P(1/3) = [0.741 \quad 0.259 \quad 0.210 \quad -0.105] \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0.505 & -0.148 \end{bmatrix} =$$

$$= [0.416 \quad 0.484]$$

и при $\tau = 2/3$.

$$P(2/3) = [0.259 \quad 0.741 \quad 0.105 \quad -0.210] \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0.505 & -0.148 \end{bmatrix} =$$

$$= [0.740 \quad 0.876].$$

Полный результат приведен в табл. 5-1. Кубический сплайн показан на рис. 5-9.

Таблица 5-1 Результаты для кубического сплайна

Сегмент	τ	$P_x(\tau)$	$P_y(\tau)$
1	1/3	0.416	0.484
	2/3	0.740	0.876
2	1/3	1.343	0.457
	2/3	1.657	-0.457
3	1/3	2.260	-0.876
	2/3	2.584	-0.484

Хотя из весовых функций (см. (5-21)) видно, что касательные векторы в концах меньше влияют на форму сплайна, чем положение концевых точек, их эффект может оказаться значительным. На рис. 5-10 изображен один симметричный сегмент сплайна с одинаковым направлением касательных векторов, но разной величины. Направления касательных векторов обозначены углом α , а их относительная величина — длиной этих векторов. Если величина значительно меньше длины хорды l , то кривая выпукла на концах и лежит внутри треугольника из хорды и касательных. При возрастании величины кривая постепенно становится вогнутой и выходит за треугольник. В этом случае при величине вектора $3/l \cos \alpha$ у кривой появляется вершина (см. рис. 5-10d). При еще больших величинах появляется петля, как видно из рис. 5-10e. Иногда для улучшения формы кривой величина вектора ограничивается длиной хорды.

5-4 НОРМАЛИЗОВАННЫЕ КУБИЧЕСКИЕ СПЛАЙНЫ

Рассмотрим другой подход к аппроксимации величин t_k -нормализацию, т. е. $0 \leq t \leq 1$ для всех сегментов.

В этом случае весовые функции (см. уравнение 5-21)¹ имеют вид:

$$F_1(t) = 2t^3 - 3t^2 + 1, \quad (5-25a)$$

$$F_2(t) = -2t^3 + 3t^2, \quad (5-25b)$$

$$F_3(t) = t^3 - 2t^2 + t, \quad (5-25c)$$

$$F_4(t) = t^3 - t^2. \quad (5-25d)$$

Матрица весовых функций для нормализованного кубического сплайна выглядит следующим образом:

$$[F] = [T][N] = [t^3 \quad t^2 \quad t \quad 1] \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (5-26)$$

¹ Данные функции являются кубическими весовыми функциями Эрмита на интервале $0 \leq t \leq 1$.

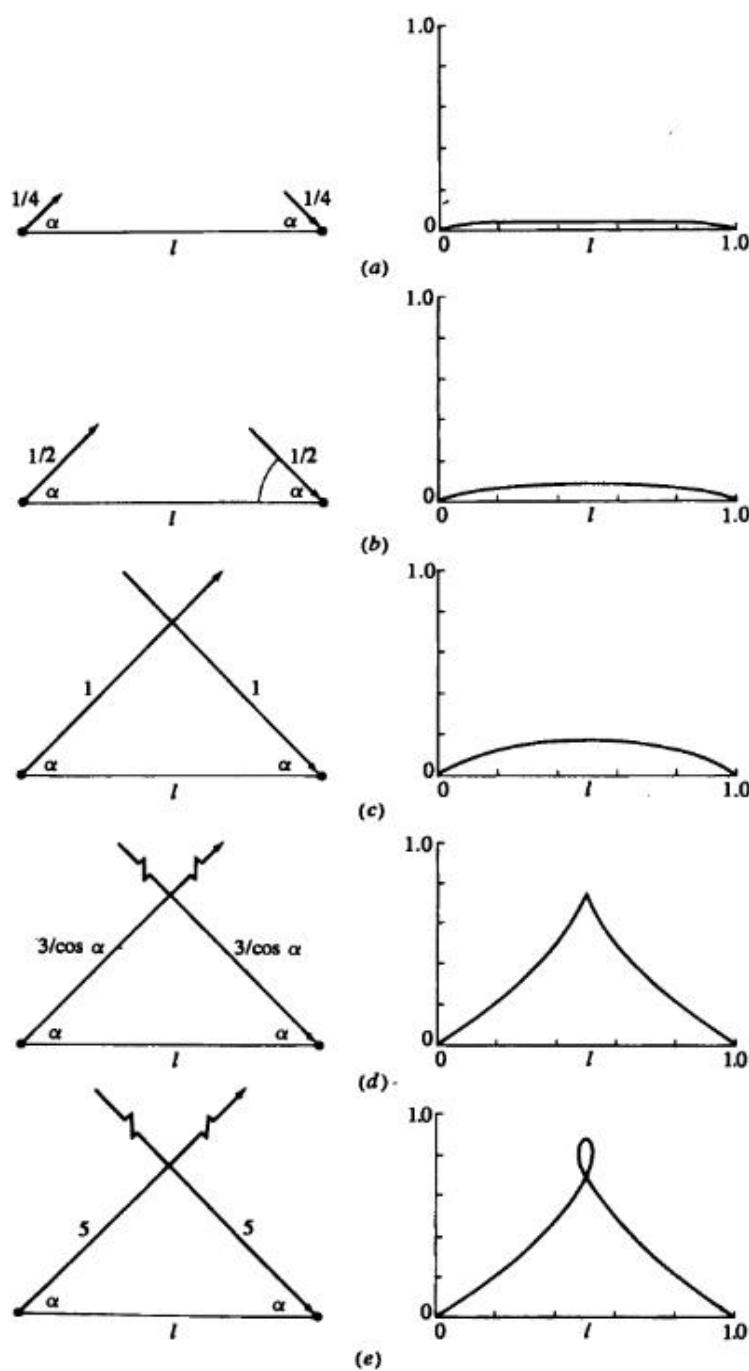


Рис. 5-10 Влияние величины касательного вектора на форму сегмента кубического сплайна, $\alpha = \pi/4$. (а) $1/4$; (б) $1/2$; (в) 1 ; (г) $3/\cos \alpha$; (д) 5 .

Матричное уравнение сегмента (уравнение (5-22)) записывается как

$$P(t) = [F][G] = [T][N][G]. \quad (5-27)$$

Заметим, что $[T]$ и $[N]$ — константы для всех сегментов сплайна. От сегмента к сегменту меняется только матрица геометрии $[G]$.

Уравнение (5-15) для определения внутренних касательных векторов, необходимых в $[G]$, приобретает вид:

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ 1 & 4 & 1 & 0 & \cdot & \cdot \\ 0 & 1 & 4 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 1 & 4 & 1 \\ \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ \vdots \\ \vdots \\ P'_n \end{bmatrix} = \begin{bmatrix} 3\{(P_3 - P_2) + (P_2 - P_1)\} \\ 3\{(P_4 - P_3) + (P_3 - P_2)\} \\ \vdots \\ \vdots \\ 3\{(P_n - P_{n-1}) + (P_{n-1} - P_{n-2})\} \end{bmatrix}. \quad (5-28)$$

Решение опять дается уравнениями (5-16), но так как $[M]$ — константа, ее нужно обратить только один раз. При большом количестве точек это значительно сокращает вычисления.

Рассмотрим пример.

Пример 5-3 Нормализованный кубический сплайн

Опять рассмотрим четыре вектор-точки на плоскости из примера 5-2, т. е. $P_1[0 \ 0]$, $P_2[1 \ 1]$, $P_3[2 \ -1]$ и $P_4[3 \ 0]$ с касательными векторами $P'_1[1 \ 1]$ и $P'_4[1 \ -1]$. Найти нормализованный кусочный кубический сплайн, проходящий через эти вектор-точки.

Величины параметров $t_2 = t_3 = t_4 = 1.0$.

Касательные векторы в точках соединения рассчитываются из уравнения (5-28), т. е.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \end{bmatrix} = \begin{bmatrix} [1 \ 1] \\ 3\{[2 \ -1] - [1 \ 1] + [1 \ 1] - [0 \ 0]\} \\ 3\{[3 \ 0] - [2 \ -1] + [2 \ -1] - [1 \ 1]\} \\ [1 \ 1] \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \\ 6 & -3 \\ 6 & -3 \\ 1 & 1 \end{bmatrix}.$$

Обращаем и умножаем:

$$\begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.267 & 0.267 & -0.067 & 0.067 \\ 0.067 & -0.067 & 0.267 & -0.267 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 6 & -3 \\ 6 & -3 \\ 1 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 \\ 1 & -0.8 \\ 1 & -0.8 \\ 1 & 1 \end{bmatrix}.$$

Отметим, что здесь внутренние касательные P'_2 , P'_3 существенно отличаются от результатов хордовой аппроксимации в примере 5-2.

Таблица 5-2 Результаты для нормализованного кубического сплайна

Сегмент	t	$P_x(t)$	$P_y(t)$
1	1/3	0.333	0.467
	2/3	0.667	0.933
2	1/3	1.33	0.422
	2/3	1.667	-0.422
3	1/3	2/333	-0.933
	2/3	2/667	-0.467

Из уравнения (5-26) матрица весовых коэффициентов для первого сегмента при $t = 1/3$ имеет вид:

$$[F] = [T][N] = [1/27 \quad 1/9 \quad 1/3 \quad 1] \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \\ = [20/27 \quad 7/27 \quad 4/27 \quad -2/27]$$

и при $t = 2/3$:

$$[F] = [7/27 \quad 20/27 \quad 2/27 \quad -4/27].$$

Точка на первом сегменте сплайна при $t = 1/3$:

$$P(t) = [F][G] = [20/27 \quad 7/27 \quad 4/27 \quad -2/27] \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & -0.8 \end{bmatrix} = \\ = [1/3 \quad 63/135] = [0.333 \quad 0.467]$$

и при $t = 2/3$:

$$P(t) = [F][G] = [7/27 \quad 20/27 \quad 2/27 \quad -4/27] \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & -0.8 \end{bmatrix} = \\ = [2/3 \quad 126/135].$$

Результаты собраны в табл. 5-2.

На рис. 5-9 изображены два сплайна — из примера 5-1 и 5-3. Кривые получились разными, т. е. параметризация влияет на результат.

Если точки из набора данных распределены неравномерно, то опыт показывает, что хордовая аппроксимация t_k дает более гладкую кривую, чем нормализация. Это видно из рис. 5-11. На рис. 5-11a показано упорядоченное множество точек, представляющее контур петушки. На рис. 5-11b точки соединены отрезками прямых. Рисунки 5-11c и d представляют нормализованную и хордовую аппроксимацию сплайна, соответственно. Обратите внимание на разницу в изображении

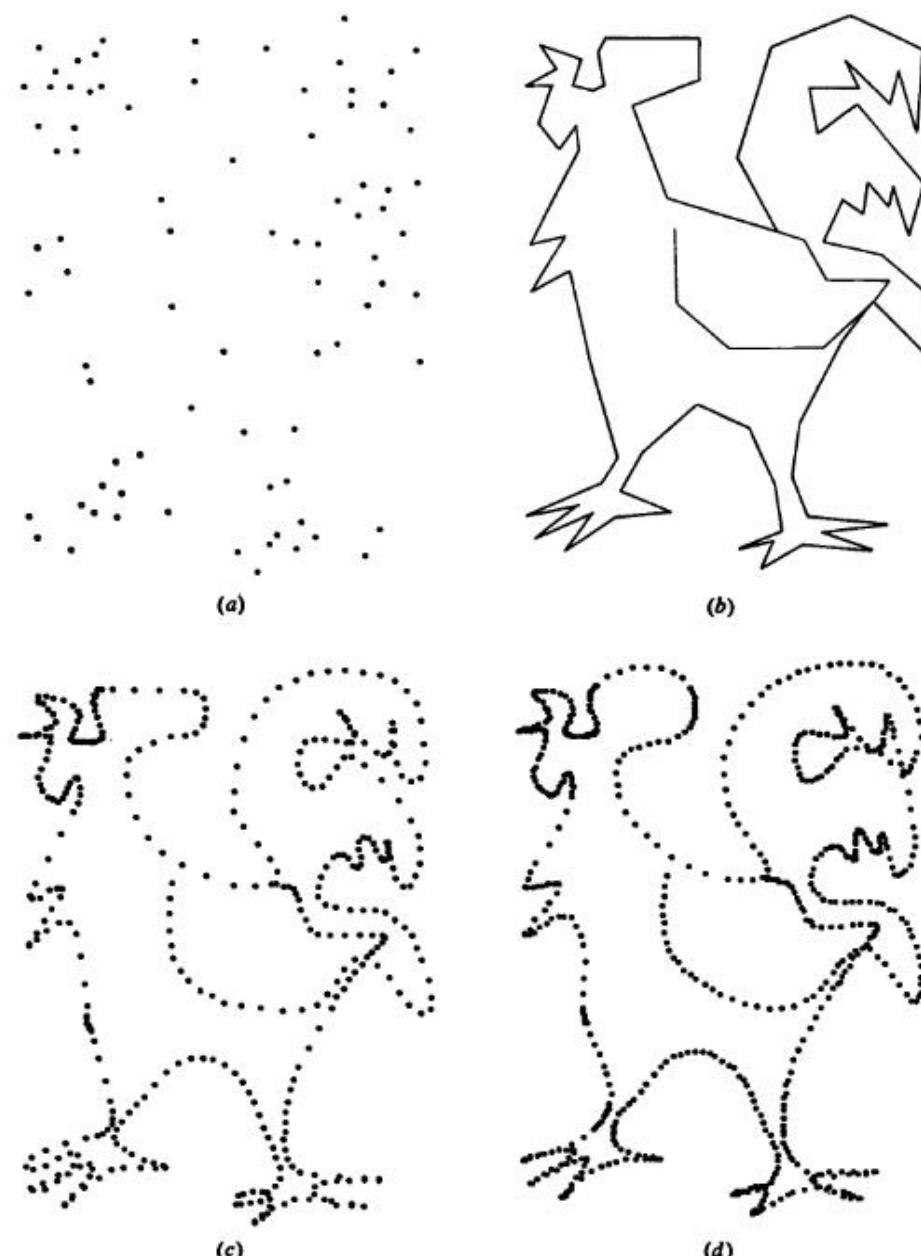


Рис. 5-11 Сравнение кубических сплайновых аппроксимаций. (a) Данные; (b) соединение прямыми; (c) нормализованная аппроксимация t_k ; (d) хордовая аппроксимация t_k .