Доказательность и эвристика при распознавании визуальных образов

В. Н. Козлов

В статье в краткой форме излагаются основные идеи, положенные в основу дискретно-геометрического подхода к распознаванию зрительных образов, и на содержательном уровне описываются полученные результаты. Приводятся некоторые соображения о роли эвристики в распознавании визуальных образов.

Ключевые слова: распознавание образов, зрительные образы, 3D-реконструкция изображений.

В этой статье сравниваются доказательные (математические) подходы к описанию и распознаванию изображений, и подходы, в основе которых лежит эвристика. При этом содержательно описывается дискретно-геометрический подход к распознаванию визуальных образов и основные его результаты. Описание разбито на 6 разделов.

1.

Говоря о распознавании изображений, нужно прежде всего четко определить, что есть изображение. Изображением здесь называется конечное множество точек в евклидовом пространстве. В частности, двумерным или плоским изображением называем конечное множество точек на плоскости.

Однако мало дать формальное определение некоторому понятию, надо еще и обосновать его соответствие содержательному аналогу этого понятия — реальному изображению. Возьмем некоторую фигуру, например, изображение буквы или цифры на плоскости. Распределим по линии фигуры более или менее равномерно некоторое количество точек, например, 30–50 точек. Если затем убрать («стереть») линии фигуры, и оставить только точки, то фигура будет явно узнаваема. Если сделать тоже самое посредством 500–800 точек, то аппроксимация точками будет мало отличаться от исходной фигуры. Тем самым «модель» фигуры в виде набора точек можно сделать сколь угодно близкой к оригиналу, вопрос лишь в том, какое количество точек мы можем себе позволить на эту аппроксимацию.

Аналогично можно представлять не только черно-белые, но и «черно-серо-белые», «полутоновые» изображения, с различными градациями «серого» цвета: в этом случае разные оттенки «серого» цвета будут представляться разной плотностью точек в разных частях изображения.

Наглядно полутоновое изображение из точек можно представить некоторым преобразованием полутонового (нецветного) пиксельного изображения, например, изображения на экране монитора. Пиксел можно считать небольшим квадратом с k возможными градациями от черного через серые до белого цвета. Представим черный пиксел k точками, равномерно распределенными по квадрату, белый пиксел — квадратом без точек, все градации между черным и белым цветом — как соответствующее число точек, распределенных по квадрату. Если это проделать с каждым пикселом, то пиксельное изображение перейдет в изображение из точек, причем зрительно это будет по прежнему восприниматься как полутоновая картина со всеми деталями.

Можно отметить, что обратная процедура — это способ превратить изображение из точек в пиксельное изображение.

В рамках таких представлений можно работать и с цветными изображениями. Это следует из того известного обстоятельства, что цветное изображение эквивалентно трем не цветным (монохроматическим), снятым в основных цветах, и образуется их наложением.

Наконец, все, что мы видим — мы видим посредством глаз. Изображение из среды проецируется на сетчатку, при этом часть чувствительных нервных окончаний (их около 130 миллионов) на сетчатке возбуждается, то есть формируется аналог точечного изображения. Именно оно представляет среду и далее анализируется мозгом.

Аналогично двумерному случаю, конечное множество точек в трехмерном пространстве может аппроксимировать с нужной степенью точностью предметы, сцены, и в целом окружающую среду. Например, некоторый реальный предмет можно представить разбитым на маленькие «кусочки», не превышающие по размеру некоторую заданную величину, и затем заменить каждый такой «кусочек» точкой в его центре. Предмет как таковой исчезнет, и останется его «модель» в виде множества точек в трехмерном пространстве. Эта модель может сколь угодно точно аппроксимировать исходный объект, если увеличивать количество точек.

Наконец, конечные множества точек в четырехмерном пространстве могут представлять трехмерные сцены в динамике. При этом одна из осей интерпретируется как ось времени.

Ясная и очевидная интерпретация для множеств точек заканчивается четырехмерным пространством. Однако можно рассматривать конечные множества точек и в пространствах большей размерности. Изображения в этом случае можно, например, трактовать в проекции на четырехмерное пространство обычным образом, а дополнительные оси пространства интерпретировать как отображающие некоторые специальные свойства изображений.

Главное в рамках рассматриваемого подхода — визуальные образы. Отметим, однако, что в эти рамки укладывается и рассмотрение звуковых образов. Например, произносимую речь можно «визуализировать», превратив посредством микрофона в «квазисинусоиду» на бумажной ленте, то есть в плоское изображение. Это изображение затем можно аппроксимировать конечным множеством точек и анализировать так же, как и любое другое.

2.

Итак, основной объект рассмотрения — конечные множества точек в евклидовых пространствах. С таким объектом можно уже работать формальными, математическими средствами, то есть доказывать теоремы, делать количественные оценки, и пр. Теоремы в этом случае — это утверждения, характеризующие свойства изображений.

На этих свойствах можно основывать алгоритмы распознавания, базирующиеся уже не на правдоподобных рассуждениях, а на точных доказательствах. Это то, что можно назвать доказательным, теоремным уровнем рассмотрения проблем распознавания изображений.

К сожалению, сейчас в исследованиях по зрительным образам превалирует эвристический подход. Этот подход основывается на здравом смысле и изобретательности, ориентированных на конкретную задачу, ее особенности. Главным является удовлетворительный в смысле практического применения результат, основной частью которого является программная реализация. Чаще всего для каждой задачи придумывается свой подход и свой способ ее решения. Тем самым, много задач — много решений. Эти решения иногда похожи, иногда сильно разнятся. В качестве примера можно вспомнить алгоритм распознавания цифр почтового индекса на конвертах. Это настолько очевидный и простой алгоритм, что его и распознаванием назвать трудно. Другой пример — программы по распознаванию рукописного текста. Задача здесь похожая, но при существенно более сложных исходных условиях.

Счастье (или несчастье) рассматриваемой тематики состоит в том, что почти все имеющиеся сейчас значимые для практики задачи могут быть решены в рамках эвристики. В мире большое число институтов, лабораторий, фирм, занимающихся созданием программ для обработки и распознавания изображений. Из чтения доступной литературы возникает ощущение, что такого рода (эвристические) решения уже давно начали повторяться (и даже, может быть, не по первому кругу). В целом все эти эвристики, существующие в мире, можно представить как гигантский совокупный комплекс (в реальности распределенный по разным местам), предоставляющий решения для широкого спектра задач из практики. Это своеобразный сборник, коллекция решений, все время пополняемая.

Однако встает вопрос: а чем предпочтительнее математический подход к распознаванию, коль скоро эвристика позволяет (действительно позволяет!) находить удовлетворительные решения практических задач? Так ли важно, чтобы у этих решений было бы еще и математическое обоснование?

Первое, что можно сказать, это то, что единое, общее, в рамках одного подхода решение целого спектра задач (если оно, конечно, существует), всегда предпочтительнее сборника разнородных решений. Стремление к такого рода общности — одна из основ научного поиска.

Главное же, что дает аналитический, математический подход к проблемам распознавания, состоит, можно полагать, в том, что возможны непредсказуемые заранее находки, неожиданные решения. В рамках эвристики мы уже в начале поиска примерно представляем себе, что хотим получить в конце. Далее в меру своей изобретательности мы подгоняем поиск под нужный «запрограммированный» ответ. У доказательного же, математического поиска своя, «внутренняя» логика, которая может вывести на совершенно неожиданные результаты. Такое неоднократно было в физике, в ее взаимодействии с математикой. Можно сослаться, например, на теории пространства-времени и их связи с геометрией и топологией, с алгеброй, и т. д.

Ниже мы приведем примеры таких неожиданностей в рамках рассматриваемой здесь модели. Они, разумеется, достаточно скромные, но, вместе с тем, их явно трудно было ожидать в начале предпринимавшегося исследования. Однако такого рода результаты возникают, и на их основе можно строить новые алгоритмы распознавания.

Было бы, однако, неправильным не отметить, что в чистом виде только теоретическими или только эвристическими подходы к распознаванию бывают далеко не всегда. Нередко в эвристических по характеру работах присутствуют элементы теоретических построений, так же как и в теоретических работах, при их программной реализации, с неизбежностью возникают элементы эвристики.

3.

Итак, два изображения A и B — это два конечных множества точек, например, две фигуры, в значительной степени повторяющие очертаниями, формой друг друга, и нам это обстоятельство нужно выяснить. Проблема, однако, в том, что априори нам неизвестно, что это за фигуры, более того, неизвестно, где у них правые-левые части, верх-низ, то есть нет ничего, кроме двух множеств точек. Как можно было бы определить схожесть по форме этих фигур? Если бы они были одинаковы, то, очевидно, это можно было бы выяснить наложением их друг на друга (например, движениями, то есть изометрическими преобразованиями) так, чтобы они полностью (поточечно) совпали. Если же фигуры не вполне одинаковы, то надо движениями расположить их одну на другой так, чтобы степень их «несовпадения», «рассогласования» была бы наименьшей из возможных. В этом случае величина этого «рассогласования» была бы количественной характеристикой несовпадения по форме.

Сделаем это требуемое следующим образом. Для каждой точки каждого из изображений A и B в качестве характеристики рассогласования этой точки со сравниваемым изображением возьмем отрезок между нею и ближайшей точкой сравниваемого изображения. Тогда характеризовать степень несовпадения в целом данных двух изображений A и B будем наибольшим из отрезков, взятых по всем точкам обоих изображений. Обозначим длину этого отрезка через $\Delta(A, B)$.

Величина $\Delta(A, B)$ характеризует взаиморасположение конкретных множеств точек A и B. Теперь будем преобразовывать A и B, например, движениями так, чтобы они как можно более совпали, то есть ищем такое взаиморасположение A и B, при котором величина $\Delta(A, B)$ наименьшая из возможных.

Нетрудно видеть, что это задача на поиск экстремума: каждое из возможных положений для A и B характеризуется вполне определенной величиной $\Delta(A, B)$; надо построить множество всех возможных взаиморасположений для A и B, порождаемых некоторыми их преобразованиями (движениями, например). Это даст множество $\{\Delta(A, B)\}$ величин $\Delta(A, B)$. Надо найти минимум на этом множестве (если он существует, конечно).

Проблема в том, что множество $\{\Delta(A, B)\}$ — бесконечное, более точно — континуальное по мощности. Таким оно будет, даже если взять в качестве преобразований для A и B наиболее простые преобразования — параллельного переноса. Естественно, проблема сохраняется, когда мы расширяем класс преобразований до изометрических — сочетаний параллельных переносов, поворотов и преобразований симметрии относительно прямой. Следующий по сложности класс — преобразования подобия, когда добавляется возможность для A и B произвольно меняться в размерах. Наконец, в случае аффинных преобразований мы добавляем ко всему перечисленному еще и возможность для A и B сжиматься и растягиваться по произвольным направлениям.

Наиболее интересен с практической точки зрения класс аффинных преобразований. Ибо в этом случае мы сравниваем «по форме» два изображения A и B безотносительно к их положению на плоскости, поворотам, изменениям в размерах, сжатиям и растяжениям. Однако, напомним, мы не можем сравнивать A и B непосредственно перебором, попробовав все варианты взаиморасположений, ибо множество $\{\Delta(A, B)\}$ (точнее, его аналог для аффинного случая) — бесконечное.

Неожиданностью здесь было то, что, как теперь доказано, существует конечное подмножество взаиморасположений для A и B, и, соответственно, конечное подмножество на $\{\Delta(A, B)\}$, на котором нужный минимум только и может достигаться. Это конечное множество взаиморасположений A и B может быть построено. Это, в свою очередь, означает, что бесконечный и потому неосуществимый перебор всех возможных вариантов взаиморасположений A и B может быть заменен проверкой лишь конечного, заранее определенного числа их взаиморасположений. Таким образом, нет необходимости пробовать по разному смещать, поворачивать изображения A и B относительно друг друга, менять их размеры, и пр. — можно сразу идти туда, где только и может быть решение.

Ясно, что на этой основе можно строить алгоритмы распознавания изображений.

4.

Одна из особенностей зрительной информации — ее огромные объемы [1, 2]. Глаз человека содержит около 130 миллионов светочувствительных элементов (палочек и колбочек). Вместе с тем распознавание изображений осуществляется иногда лишь за доли секунды. Вряд ли это можно сделать, используя целиком всю информацию на сетчатке глаза. Должно существовать некоторое «сито» и для изображений из среды, и для визуальной информации из памяти при оперировании с ними. Парадокс, однако, состоит в том, что попытки до собственно распознавания выделить на изображениях «опорные точки», «важные детали» и пр. (как правило, средствами эвристики), есть тоже распознавание, то есть возникает до некоторой степени замкнутый круг. Ясно, что нужно избегать этого.

Трудно рассчитывать эту задачу — построения «сита» для изображений — решить некоторым единовременным усилием. В рамках описываемой модели предлагается лишь некоторый шаг в ее решении.

Зададимся некоторым числом r ($r \geqslant 0$). Выберем на изображении A подмножество A^r точек из A с тем условием, чтобы каждая точка из A находилась бы на расстоянии, не большем r, от какой-либо точки из A^r. Ясно, что A^r можно трактовать как подизображение изображения A, в частном случае A^r может совпадать с A. Изображение A^r можно интерпретировать как покрытие A кругами радиуса r с центрами в точках из A^r, причем такое, что каждая точка исходного изображения попадает хотя бы в один из этих кругов.

Содержательно A^r трактуется как «огрубление» изображения A, устранение на нем излишних деталей и подробностей, поэтому A^r называем эскизом изображения A. При заданном параметре r существует некоторое множество $\{A^r\} = \{A^r_1, \ldots, A^r_k\}$ эскизов, причем исходное изображение A входит в это множество. Однако интерпретировать как «огрубление» A можно, очевидно, лишь те эскизы из A^r, которые состоят из меньшего числа точек, чем исходное изображение A. В частности, в $\{A^r\}$ есть изображение (может быть, не одно) с наименьшим среди $A^r_1, \ldots, A^r_k$ числом точек. Такое изображение назовем остовом изображения A.

Можно, очевидно, говорить об иерархии остовов, определяемой различными значениями r. Остовов (как и эскизов) — конечное множество, и наименьший по числу точек остов всегда (и для всех изображений) состоит из одной точки, наибольший — совпадает с исходным изображением A.

При оценке похожести двух изображений достаточно, может быть, оценить их сходство в целом, в главных чертах, без деталей. Такого рода интуитивные соображения неявно предполагают, что вместо собственно изображений будут использованы другие изображения, полученные из исходных устранением излишних деталей, то есть в нашем случае — эскизы. Оказалось возможным теоремным образом связать схожесть между эскизами со схожестью между оригиналами. Это дает возможность, имея дело с эскизами, делать оценки для схожести исходных изображений. Распознавание при таком подходе перестает прямо зависеть от количества точек в исходных изображениях, и даже появляется потенциальная возможность рассматривать изображения, бесконечные по числу составляющих их точек («непрерывные» изображения).

5.

Поместим перед глазом плоское точечное изображение некоторой фигуры, например, цифры. В реальном эксперименте такую фигуру можно изобразить точками черной краски на плоской прозрачной стеклянной пластинке. Сетчатку полагаем плоской. Изображение помещается перед глазом на некотором расстоянии так, чтобы его плоскость была бы параллельна плоскости сетчатки.

На сетчатке сформируется совокупность возбужденных рецепторов (или, в нашей интерпретации, конечное множество точек плоскости). Эта совокупность — и ничего более — есть тот первичный материал, который будет далее анализироваться мозгом в процессе распознавания. Если теперь изображение перед глазом сдвинуть (то есть сместить его, сохраняя неизменным расстояние до глаза и сохраняя плоскость изображения параллельной плоскости сетчатки), то и изображение на сетчатке сдвинется (преобразование в виде параллельного переноса). При этом изображение на сетчатке останется, в сущности, прежним, однако первичный материал — совокупность возбужденных рецепторов — может, очевидно, существенно поменяться. Понимание того, что объект не изменился, можно, очевидно, основывать на таком описании (кодировке) изображений, которое не меняется при параллельных переносах.

Мы рассмотрим далее разные варианты перемещений фигуры перед глазом и соответствующих преобразований изображения на сетчатке. Набор перемещений, который мы рассмотрим, будет полон в том смысле, что любое перемещение плоского изображения перед глазом можно рассматривать как сочетание этих перемещений. Для каждого случая будем строить кодировку изображения, инвариантную к возникающим преобразованиям на сетчатке.

Случай 1. Этот случай соответствует, в сущности, рассмотрению того, что выше было названо первичным материалом. Изображение A представляется совокупностью конкретных возбужденных рецепторов. Положение каждого такого рецептора (точки) известно. Его можно задавать по разному. Будем полагать, что на сетчатке введена декартова система координат и положение каждого рецептора определяется его координатами. Отметим, что можно было бы взять и другие системы координат — суть дела это меняет незначительно.

Перенумеруем точки изображения A некоторым образом так, чтобы номера были попарно различны. Кодом K^1 изображения A (обозначение K^1_A) назовем пару множеств $\langle M_A, T_A\rangle$. Здесь M_A — множество номеров точек изображения A, T_A — множество координат $\{(x,y)_i\}$ точек с указанием их номера, то есть, например, $(5,3)_7$ означает, что у точки с номером 7 координатами является пара $(5,3)$.

Перенумеровав по иному множество точек изображения A, получим другую пару $\langle M_A, T_A\rangle$. Будем, однако рассматривать кодировку с точностью до перенумерации и называть изображения с кодами, отличающимися только нумерацией, эквивалентными (в смысле случая 1). Таким образом, если имеются изображение A с кодом $\langle M_A, T_A\rangle$ и B с кодом $\langle M_B, T_B\rangle$, то назовем A и B эквивалентными (в смысле случая 1), если существует такая нумерация точек изображения A, при которой его код есть $\langle \tilde{M}_A, \tilde{T}_A\rangle$, и такая нумерация для B, при которой его код есть $\langle \tilde{M}_B, \tilde{T}_B\rangle$, и при этом $\tilde{M}_A = \tilde{M}_B$ и $\tilde{T}_A = \tilde{T}_B$.

Случай 2. Пусть изображение перед глазом некоторым образом смещается и поворачивается. Пусть при этом плоскости изображения и сетчатки остаются параллельными (или, если они не параллельны, то угол между этими плоскостями остается неизменным) и расстояние между плоскостями не меняется. Преобразования изображения на сетчатке сводятся, очевидно, к тому, что оно смещается и поворачивается — других изменений нет. Ясно, что при этом меняется и координатный код изображения, поскольку меняются и координаты составляющих изображение точек.

При сдвиге и повороте меняется положение изображения на сетчатке, другими словами, положение изображения по отношению к осям системы координат. Взаиморасположение же точек изображения не меняется, что означает сохранение расстояний между ними. Именно сохранение расстояний между точками при сдвиге и повороте положим в основу нового кода K_A^2. В коде K_A^2, как и ранее, M_A есть множество попарно отличающихся, а в остальном произвольных, номеров точек изображения. Множество T_A составляют все числа $r(m, n)$ (m и n — номера точек из M_A), являющиеся расстояниями между точками изображения, с указанием того, расстоянием между какими точками является данное число. Аналогично случаю 1 можно определить одинаковые с точностью до перенумерации точек коды. Изображения с такими кодами тоже назовем эквивалентными (в смысле случая 2). Можно показать, что два изображения эквивалентны тогда и только тогда, кода одно может быть получено из другого комбинацией сдвига, поворота и преобразования симметрии.

Рассмотрим теперь, как соотносятся коды K_A^1 и K_A^2 и чем объясняется различие их свойств. Задавая изображение совокупностью координат всех его точек (код K_A^1), мы, в сущности, задаем нечто большее, чем собственно изображение — неявным образом в таком задании присутствует и внешняя по отношению к изображению система отсчета, то есть система координат с ее осями. Эта система отсчета в реальности должна быть «привязана» к каким-то дополнительным внешним точкам, например, к краям сетчатки. Действительно, посмотрим на код K_A^1 как на некий вариант кода K_A^2, то есть будем считать, что каждая точка в K_A^1 тоже задается совокупностью расстояний до других точек. В таком случае этими другими точками для произвольной точки a изображения будут точки x_a и y_a на осях координат, являющиеся проекциями точки a на оси. Код K_A^1 как бы предполагает наличие, помимо точек собственно изображения, еще и точек x_a и y_a для каждой точки a. Если присоединить эти точки к изображению, то, очевидно, код K_A^2 уже не будет инвариантным к сдвигу и повороту первоначального изображения, поскольку при этом меняются расстояния от точек первоначального изображения до добавленных точек.

Случай 3. Пусть изображение перед глазом приближается или удаляется с сохранением параллельными плоскости изображения и плоскости сетчатки (или, если они не параллельны, с сохранением неизменным угла между ними). Изображение на сетчатке при этом увеличивается или уменьшается в размерах с сохранением подобия.

Коды K_A^1 и K_A^2 меняются, поскольку меняются и координаты точек изображения на сетчатке, и расстояния между точками. Соотносительные же размеры частей изображения при увеличении или уменьшении с сохранением подобия не меняются. Это и положим в основу кода K_A^3. Так же, как и в предыдущих случаях, перенумеруем множество точек изображения A и обозначим множество номеров через M_A. Далее зададим множество численных значений отношений вида $r(m,n)/r(p,q)$, где $r(m,n)$ и $r(p,q)$ — расстояния между точками с номерами соответственно m и n, p и q. Здесь m,n,p,q — номера из M_A, $m \neq n$, $p \neq q$. Для каждого числа из T_A полагаем известной соответствующую ему четверку номеров. Код K_A^3 есть пара $\langle M_A, T_A \rangle$.

Аналогично тому, как это рассматривалось для предыдущих случаев, можно определить одинаковые с точностью до перенумерации точек коды (эквивалентные в смысле случая 3). Можно показать, что два изображения эквивалентны в том и только в том случае, если на плоскости одно может получено из другого сдвигом, поворотом, изменением в размерах (с сохранением подобия), преобразованием симметрии или их комбинацией (то есть подобными преобразованиями).

Рассмотрим теперь, как соотносятся коды K_A^2 и K_A^3 и чем объясняется различие их свойств. По коду K_A^2 можно, очевидно, построить код K_A^3. Обратное неверно. Код K_A^2 определялся таким образом, чтобы в описании изображения не участвовала бы «внешняя» по отношению к изображению система координат. Однако нечто от этой внешней системы в определении кода все же осталось, а именно — единица измерения расстояний между точками (масштабная единица). Она, эта единица, предполагается для кода K_A^2 априори заданной. Когда же мы берем отношение $r(m,n)/r(p,q)$ (в коде K_A^3), эта единица измерения устраняется. Действительно, величина отношения $r(m,n)/r(p,q)$ будет одной и той же вне зависимости от того, в каких единицах измеряются расстояния $r(m,n)$ и $r(p,q)$. Отсюда и возникает возможность посредством K_A^3 описывать изображение безотносительно к его размерам.

Можно и несколько по иному интерпретировать код K_A^3. Возьмем в качестве единицы измерения расстояние между какой-либо парой точек на самом изображении. Пусть это будут точки с номерами p и q. Теперь для того, чтобы для произвольных точек m и n изображения получить расстояние между ними, выраженное в единицах, являющихся расстоянием между точками p и q, нужно, очевидно, проделать следующее. Взять произвольную единицу измерения и с ее помощью получить расстояния $r(m,n)$ и $r(p,q)$. Затем разделить $r(m,n)$ на $r(p,q)$, то есть получить $r(m,n)/r(p,q)$. Это и будет искомым числом. Если теперь поочередно считать единицей измерения расстояние между каждой парой точек в изображении, то мы и придем ко множеству T_A для кода K_A^3.

Случай 4. Проведем в плоскости изображения перед глазом прямую. Повернем изображение вокруг прямой на некоторый угол. Изображение на сетчатке сожмется в направлении, перпендикулярном к прямой.

Если в исходном положении плоскость изображения не параллельна плоскости сетчатки и затем поворачивается, делаясь параллельной, то, очевидно, изображение на сетчатке при этом будет растягиваться в направлении, перпендикулярном к оси поворота.

Подчеркнем, что к сжатию или растяжению преобразование изображения на сетчатке сводится в том случае, когда несущественна разница в расстояниях до глаза от разных частей изображения, возникающая после поворота вокруг оси. В целом эта особенность возникает как часть более широкой проблемы, отмеченной в исследованиях по машинному зрению (например, в [3, 4]). Проекции тела на сетчатку рассматриваются обычно как параллельные проекции (с добавление возможности увеличения-уменьшения в размерах). Глазом же, можно считать с большим основанием, осуществляется центральная проекция. Если, однако, расстояние до объекта велико по сравнению с размерами самого объекта, то различия между центральной и параллельной проекциями невелики и ими можно пренебречь. Мы считаем, что находимся в границах применимости именно этого случая.

Коды K_A^1, K_A^2, K_A^3 изображения при сжатии и растяжении, очевидно, меняются. Так, например, код K_A^3 меняется потому, что меняются соотносительные размеры частей изображения (не сохраняется подобие).

Мы последовательно устраняли в описаниях изображения (кодах) зависимость от внешней системы координат и, казалось бы, от нее ничего не осталось. Последнее, что мы сделали (код K_A^3) — устранили единицу измерения расстояний. Однако, несмотря на то, что самой единицы уже нет, нечто от нее осталось — предположение, что она одна и та же по всем направлениям на плоскости (то есть, в сущности, предположение об изотропности пространства). Откажемся теперь и от этого предположения.

Как и ранее, код изображения A — пара $\langle M_A, T_A\rangle$, где M_A — множество номеров точек. Множество T_A состоит из чисел с индексами вида $\rho_{mnu,ksp}$. Здесь $\rho_{mnu,ksp}$ получается отношением площадей треугольников с вершинами в точках изображения с номерами соответственно m, n, u и k, s, p (естественно при условии, что площадь треугольника с вершинами в точках k, s, p не равна нулю). Такие числа ρ строятся для всех пар треугольников. Подчеркнем, что только на этапе формирования числа $\rho_{mnu,ksp}$ мы говорим о треугольниках и о площадях — далее этой информации нет и T_A состоит только из чисел с индексами.

Доказано, что коды двух изображений одинаковы (с точностью до перенумерации точек) тогда и только тогда, когда эти изображения переводятся одно в другое аффинным преобразованием. Тем самым, этот код задает изображение с точностью до аффинных преобразований. Это означает, что по коду изображение можно восстановить по разному, но все варианты восстановления будут отличаться друг от друга только аффинными преобразованиями.

Аналогичным свойством обладает код $\langle M_A, T_A\rangle$ для точек в трехмерном пространстве (трехмерные изображения или тела). Различие с двумерным случаем состоит только в том, что числу ρ приписаны две четверки индексов m, n, u, v и k, s, p, q и само число получается отношением объемов тетраэдров с вершинами в соответствующих точках. Этот код тоже описывает тело с точностью до аффинных преобразований.

Доказано, что аналогичные коды существуют и для конечных множеств точек в пространствах размерности n, где $n > 3$.

Показано, что существует обобщение кода на случай проективных преобразований изображений.

6.

Пусть теперь есть некоторое тело Q (трехмерное изображение) и двумерное изображение S. Вопрос, который нас интересует, может быть поставлен так: как соотносятся Q и S, если известно, что S является проекцией тела Q на некоторую плоскость.

Эта задача является основой для восстановления трехмерного изображения по его плоским проекциям, что в свою очередь есть важнейшая составляющая моделей стереовосприятия для роботов, живых организмов и процедур построения изображений в томографии.

Вкратце обычную схему восстановления тела по проекциям можно описать следующим образом. Точка m тела Q проецируется на плоские сетчатки S_1 и S_2 и ее проекции есть соответственно m_1 и m_2. Если известно, например, положение m_1 и m_2 на сетчатках, расстояние между сетчатками, и направления лучей проекции от m к m_1 и m_2, то тривиальным образом по этой информации восстанавливается точка m. Проделывая это для каждой точки тела Q, мы восстанавливаем тело. Проекции S_1 и S_2 тела на две сетчатки несколько разные за счет того, что каждый глаз «видит» тело под своим углом зрения, в своем ракурсе. Именно этой разностью и обеспечивается возникновение стереоскопического эффекта.

Как отмечается в руководствах по компьютерному стереозрению, главной здесь является проблема нахождения соответствующих друг другу точек на двух проекциях. Выше мы говорили о двух соответствующих точках m_1 и m_2 на S_1 и S_2. Когда таких точек много, то неясно, какую из них на одной проекции сопоставлять данной точке на другой. Это нельзя сделать, например, простым наложением изображений на сетчатках друг на друга, поскольку эти изображения разные за счет разных ракурсов. Предполагать, что уже «распознано», какие части изображений соответствуют друг другу, и на этой основе сопоставлять точки проекций тоже нельзя, так как задача распознавания предполагается решаемой на более поздних этапах и ее решение отчасти должно основываться на результатах восстановления трехмерного изображения.

Дополнительные трудности возникают при попытках объяснить этой схемой механизмы стереоскопического зрения в живых организмах. Действительно, для того, чтобы знать расстояния между соответствующими точками на двух проекциях, нужно, как минимум, иметь известным расстояние между двумя сетчатками. Неясно, на каком основании мы можем считать это расстояние априори известным. Внутри одного вида расстояния между глазами у разных особей несколько разнятся, не говоря уже об особях разных видов. На протяжении жизни, вследствие изменений в размерах тела, это расстояние тоже меняется. Трудно поэтому считать его жестко генетически обусловленным и на этом основании известным. Если же предполагать, что это расстояние становится известным из индивидуального опыта, то надо каким-то образом объяснить механизм получения такого рода информации. Кроме того, если объект находится непосредственно перед глазами, то изображения на сетчатках, можно считать, одного размера. Однако если объект находится сбоку и глаза повернуты на объект, то сетчатки, очевидно, находятся на несколько разных расстояниях до объекта. Это приводит к некоторой разнице в размерах изображений на сетчатках.

В целом рассмотрения такого рода приводят к следующей задаче. Имеются плоские проекции S_1 и S_2 тела Q (проекции на сетчатки). Однако в нашем распоряжении есть только плоские изображения $\tilde{S}_1$ и $\tilde{S}_2$, полученные из соответственно S_1 и S_2 произвольными аффинными преобразованиями (какими — неизвестно). Кроме того, поточечное соответствие между $\tilde{S}_1$ и $\tilde{S}_2$ не задано. Надо восстановить тело Q.

Ясно, что в такой постановке снимается и вопрос о расстоянии между сетчатками, и о направлении лучей проекции, которыми получены точки на S_1 и S_2.

Оказывается, в такой постановке задача решаема: по $\tilde{S}_1$ и $\tilde{S}_2$ тело Q восстанавливается с точностью до аффинных преобразований, то есть строится тело $\tilde{Q}$, отличающееся от исходного Q некоторым аффинным преобразованием. При этом определяется и поточечное соответствие между $\tilde{S}_1$ и $\tilde{S}_2$. Это не эмпирический результат — соответствующая теорема доказывается.

Поскольку $\tilde{S}_1$, $\tilde{S}_2$, $\tilde{Q}$ отличаются от соответственно S_1, S_2, Q аффинными преобразованиями, то описанное можно представлять так, что по кодам плоских изображений S_1 и S_2 (задающим эти изображения с точностью до аффинных преобразований), восстанавливается код трехмерного изображения Q (задающий его тоже с точностью до аффинных преобразований).

К настоящему времени есть несколько компьютерных реализаций изложенного подхода.

1) Распознавание произвольных черно-белых фигур: цифр, букв, подписей, иероглифов, рисунков, и пр. На экране «мышкой» рисуются образцы фигур, которые нужно распознавать, они отправляются в память. Распознаваемый объект тоже рисуется на экране. Распознавание состоит в выборе того объекта из памяти, который лучше всего «натягивается» на объект на экране методами, описанными кратко в разделах 3 и 4. Это наилучшее «натяжение» выводится на экран и является ответом. Распознавание не зависит от размеров, ориентации, положения на плоскости, сжатий-растяжений и локальных особенностей сравниваемых фигур.

Есть более сложный вариант программы, когда на экране — несколько фигур, возможно пересекающихся, наложенных друг на друга. Образцы из памяти должны сами «найти» на экране те фигуры, на которые они лучше всего «натянутся».

2) Восстановление трехмерных изображений по стереопарам (стереофотографиям). Исходными являются две фотографии объекта или сцены в двух несколько разных ракурсах (обычные стереопары). Методами, изложенными кратко в разделе 6, восстанавливается трехмерное изображение («виртуальное», в памяти компьютера). Это дает возможность получать новые проекции, то есть объект в новом ракурсе (фактически новые изображения объекта). Если это делать последовательно, то на экране объект или сцена поворачиваются так, как если бы мы перемещались относительно них (конечно, в опреде-

ленных пределах: восстанавливаются только те точки сцены, которые присутствуют на обеих исходных проекциях).

3) Распознавание мелодий. Мелодии, заносимые в память и распознаваемые — только из «чистых», «синусоидальных» звуков (они сравнительно легко создаются непосредственно на компьютере). Такие мелодии легко «визуализируются» в виде двумерной картинки. Распознаваемая мелодия может отличаться от записанной в память громкостью, темпом, тональностью.

Список литературы

[1] Глезер В. Д. Зрение и мышление. Спб.: Наука, 1993.

[2] Хьюбел Д. Глаз, мозг, зрение. М.: Мир, 1990.

[3] Хорн Б. К. П. Зрение роботов. М.: Мир, 1989.

[4] Фу К., Гонсалес Р., Ли К. Робототехника. М.: Мир, 1989.

[5] Козлов В. Н. Введение в математическую теорию зрительного восприятия. М.: Издательство Центра прикладных исследований при механико-математическом факультете МГУ, 2007.