Об определении понятия визуального образа

В. Н. Козлов[1]

Работа относится к области теоретических аспектов распознавания визуальных образов. Изображения трактуются как конечные множества точек в евклидовых пространствах разной размерности. Так действительно можно представлять реальные изображения. Формального определения визуального образа пока нет. В статье делается первое приближение к этому определению.

Ключевые слова: распознавание образов, визуальные образы, изображения.

Изображением называем конечное (непустое) множество точек в евклидовых пространствах разной размерности. В частности, двумерное изображение — конечное множество точек на плоскости. Обосновываем это тем, что любую фигуру можно «аппроксимировать» конечным множеством точек (рис. 1), которые уже сами по себе делают фигуру вполне узнаваемой. Если точек много, то такая совокупность точек практически неотличима от исходной фигуры. Так же можно представлять и полутоновые, черно-бело-серые изображения, при этом разная плотность точек в разных частях изображения дает разные оттенки «серого цвета». Как известно, цветное изображение можно представлять как наложение трех монохроматических (аналогов черно-бело-серых) изображений. Это означает, что совокупностями точек можно представлять и цветные изображения. Трехмерные изображения – точки в трехмерном евклидовом пространстве. Соотнесение между трехмерным изображением и двумерными, являющимися его проекциями, приводит к задачам восстановления тел по плоским проекциям и смежным задачам в приложениях. Наконец, трехмерный мир в динамике можно рассматривать как четырехмерное изображение (последовательность трехмерных сцен).

Далее рассматриваются двумерные изображения, но сказанное несложно обобщается и на случаи большей размерности.

Среда S у нас — произвольное изображение, состоящее из $N (N \geq 1)$ точек. Частью среды называем любое непустое подмножество ее точек. Частей среды — $\left(2^N - 1\right)$. Множество частей (т.е. изображений) обозначим через S^+. Образ из S будем трактовать как некоторую группу изображений из S^+, т.е. задавать его перечислением всех возможных приме-

---

[1] *Козлов Вадим Никитович* — профессор каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: vnkozlov@mail.ru.

Kozlov Vadim Nikitovich — professor, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Mathematical Theory of Intellectual Systems.

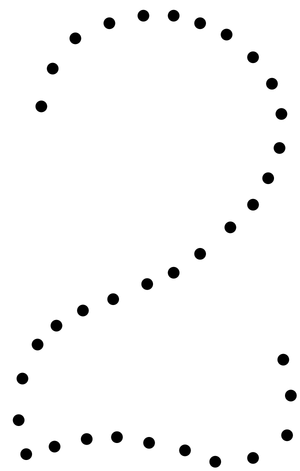

Рис. 1.

ров изображений образа в данной среде S. Обозначим множество всех возможных групп через $S^\star$. Разумеется, далеко не все из этих групп — образы в содержательном понимании. Тем самым, задача состоит в том, чтобы «ужать» множество $S^\star$ до совокупности групп таких, которые уже можно с приемлемой степенью убедительности трактовать как образы. В этом и состоит подход к определению понятия образа в данной работе.

Приближений к понятию образа предполагается несколько, в этой статье описано одно (первое приближение).

Введем сквозной для последующего изложения пример: представим для наглядности среду S как «хаос» на плоскости разных фигур — цифр, букв и пр. — разных по размерам, ориентации, пересекающихся, имеющих общие части и пр. Тогда среди групп множества $S^\star$ будут как «осмысленные», т. е. группы например «двоек», или «троек», так и «бессмысленные» — сочетания одновременно и «двоек», и «троек», их частей, и многого другого. Вот эти бессмысленные сочетания и надо отсеять, основываясь на некоторых далее вводимых принципах.

Для дальнейшего нам понадобятся некоторые понятия и ранее опубликованные результаты [1, 2, 3]. Содержательный смысл их в том, чтобы доказательным образом обеспечить наложение одного изображения на другое аффинными преобразованиями так, чтобы различие между изображениями было бы минимальным из возможных.

Пусть изображение A состоит из точек $a_1, ..., a_n$, изображение B — из точек $b_1, ..., b_n$, ψ — одно из возможных взаимно однозначных соответствий между точками изображений A и B, которым точке a_i из A сопоставляется точка $b_{\psi(i)}$ из B $(i = 1, ..., n)$. Обозначим через $B^\star$ множество всех изображений, получаемых из B аффинными преобразованиями. Полагаем, что на B' из $B^{\star}$ сохраняется нумерация, порожденная изображением B, т.е. через b'_i на B' обозначается точка, в которую переходит при соответствующем преобразовании точка b_i из B. Точки a_i и $b'_{\psi(i)}$ называем соответствующими, соответствующими называем и отрезки $(a_i a_j)$ и $\left(b'_{\psi(i)} b'_{\psi(j)}\right)$.

Зададимся некоторым положительным числом ϵ. Обозначим через $\{B\}^{\epsilon}$ множество всех таких изображений B' из $B^{\star}$, для которых длина каждого отрезка $(b_i b'_i)$ $(i = 1, ..., n)$ не больше ϵ. Преобразования, переводящие изображения из $\{B\}^{\epsilon}$ друг в друга, назовем ϵ-аффинными. Содержательно их можно трактовать как ограниченные, локальные аффинные преобразования для B.

Дадим важное определение искомого (или оптимального) взаиморасположения: через $l_A(B')$ обозначим длину наибольшего из отрезков $\left(a_i b'_{\psi(i)}\right)$ $(i = 1, ..., n)$. Рассмотрим B_0 — некоторое изображение из $B^{\star}$, и ψ_0 — одно из взаимно однозначных соответствий между точками изображений A и B. Пусть существует такое ϵ_1, что для всех B' из $\{B_0\}^{\epsilon_1}$ и при всех биекциях ψ минимум величин $l_A(B')$ достигается на изображении B_0 и при биекции ψ_0. Пусть существует такое ϵ_2, что для всякой пары изображений (A', B'_0), получаемой ϵ_2-аффинными преобразованиями пары (A, B_0) как целого, выполняется аналогичное свойство: для всех B'' из $\{B'_0\}^{\epsilon_1}$ и при всех биекциях ψ минимум величин $l_{A'}(B'')$ достигается на изображении B'_0 и при биекции ψ_0. Тогда B_0 называем искомым для изображения A (и взаиморасположение A и B_0 искомым), биекцию ψ_0 — искомым соответствием между точками в A и B.

Что есть в содержательной интерпретации искомое (оптимальное) расположение изображения B на A, например, в случае, если A «эталонное» изображение «двойки», а B — тоже «двойка», но несколько искаженная по форме? Тогда B_0 — расположенная аффинными преобразованиями на A искаженная «двойка», причем расположенная так, чтобы, несмотря на исходные искажения, максимально повторять своей формой форму неискаженной A, при этом — безотносительно к размерам, ориентациям и сжатиям-растяжениям исходной фигуры B. Параметр $l_A(B')$ и служит мерой несовпадения форм фигур. Для двух двоек он, предполагается, будет существенно меньше, чем, скажем, для «двойки» и «четверки».

Ясно, что перебором оптимальное взаиморасположение двух фигур не найти — перебор будет бесконечным.

Из полученных ранее результатов [1] следуют некоторые необходимые условия, которым должны удовлетворять изображения в паре (A, B_0). Они коротко состоят в следующем. Пусть биекцией ψ точке a_i

из A сопоставляется точка $b'_{\psi(i)}$ из B $(i = 1, ..., n)$. Возьмем на плоскости произвольную точку O и параллельными переносами отрезка $\left(a_i, b'_{\psi(i)}\right)$ совместим точку a_i с точкой O. Точку, в которую перейдет при этом $b'_{\psi(i)}$, обозначим через $c_{i\psi(i)}$ и назовем порожденной парой соответствующих точек a_i и $b'_{\psi(i)}$. Изображение из точек O и $c_{i\psi(i)}$ $(k = 1, ..., n)$ называем характеристическим, O — центр характеристического изображения, $c_{i\psi(i)}$ — точки ядра.

Окружность наименьшего по радиусу круга, включающего все точки ядра, называем ключевой. Доказано [1], что для пары (A, B_0) центр характеристического изображения с необходимостью должен совпадать с центром ключевой окружности.

Назовем изображение B' из $B^\star$ согласованным с A, если существуют в B' два непараллельных отрезка $(b'_1 b'_2)$ и $(b'_3 b'_4)$, равные, параллельные и однонаправленные с соответствующими отрезками $(a_1 a_2)$ и $(a_3 a_4)$ в A. Параллельные отрезки, например, $(a_1 a_2)$ и $(b'_1 b'_2)$ называем однонаправленными, если, при условии, что в $(a_1 a_2)$ слева направо сначала идет точка a_1, а затем a_2, то и в отрезке $(b'_1 b'_2)$ слева направо сначала идет точка b'_1, затем b'_2.

Доказано [1], что если B_0 искомое изображение для A, то B_0 согласовано с A.

Итак, имеем теперь два необходимых условия для пары (A, B_0): искомое изображение B_0 должно быть согласовано с A, и центр характеристического изображения пары (A, B_0) должен совпадать с центром ключевой окружности. Сочетание этих двух условий позволяет вычленить из $B^\star$ конечное подмножество изображений, среди которых только и может находиться искомое изображение B_0. Здесь существенна именно конечность, поскольку этим потенциально бесконечный перебор при поиске B_0 сводится к конечному.

Отметим, что в частном случае A может быть, конечно, и аффинно эквивалентным с B, и тогда искомое взаиморасположение для них состоит в том, что они просто совпадают.

Дано изображение X. Обозначим через $W(X)$ выпуклую оболочку для X. Известно [4, 5], что для выпуклого множества $W(X)$ существует единственный наименьший по площади эллипс, его вмещающий. Центр этого эллипса назовем центром изображения X, а длину большей оси — размером. Множество всех частей изображения X обозначим через X^+.

Два изображения A и C называем непересекающимися, если $W(A)$ и $W(C)$ не имеют общих точек. Рассматриваем покрытие P_X (такое всегда есть) изображения X непересекающимися частями из X^+. Части называем также кусками изображения X. Размером покрытия называем размер наибольшего куска в нем. Ясно, что возможных покрытий — конечное множество.

Уместно сделать следующее пояснение к дальнейшему. Построение модели — это всегда нечто такое, что первоначально возникает главным образом на основе интуиции, правдоподобных рассуждений. А уже затем на этой первоначальной основе строится совокупность определений (формальных, или, для начала, полуформальных), позволяющих проводить доказательные рассуждения, т.е. получать утверждения (теоремы) о свойствах модели. Эвристика, в немалой степени присутствующая в распознающих системах — это тоже правдоподобные рассуждения. Но она только ими и ограничивается, не переходя к доказательным построениям.

Далее представлены два пункта содержательных соображений, на которых в значительной мере далее строится модель.

О понятии остова для отдельного изображения и для группы изображений. Предположим в качестве гипотезы, что для группы похожих изображений во всех них есть нечто общее, объединяющее эти изображения, и что мы назовем их остовом. Опишем возникновение этого понятия на содержательном уровне в нескольких приближениях, начиная с простого и довольно очевидного.

Представим «двойку» в виде совокупности точек с кругами (рис. 2). Это основа, базовое изображение A. Трактуем его так: исходная совокупность точек разбита на куски, причем такие, что тот единственный и наименьший по площади эллипс, который вмещает кусок, есть круг, причем все круги — одного радиуса, и их центры и образуют «двойку», которую мы видим на рисунке. Это несколько искусственный пример, потому что если реальную «двойку» разбить на куски, то минимальные по площади эллипсы этих кусков вовсе не обязательно будут именно кругами. Круги могут — немного — пересекаться. Вписанные же в круги куски (они на рисунке не представлены), как и следует из определений выше, пересекаться не могут. Предполагается, что при фиксированном радиусе количество кругов наименьшее из возможных.

Заменим теперь содержащееся в каждом круге множество точек (на рис. 2, напомним, они не показаны) на другое, но с тем же условием: выпуклая оболочка на этом множестве порождает кусок, вписанный в круг, причем центр куска совпадает с центром круга, куски из разных кругов, естественно, не пересекаются. Это второе изображение, ясно, отличается от первого. Аналогично можно породить целый класс изображений, обозначим их $A_1, ..., A_k$ (потенциально бесконечный), они разные (и разнообразные), поскольку в круги можно «вставлять» разные изображения, даже, например, те же «двойки», только маленькие по размерам. Изображение из центров кругов для каждого из изображений можно на-

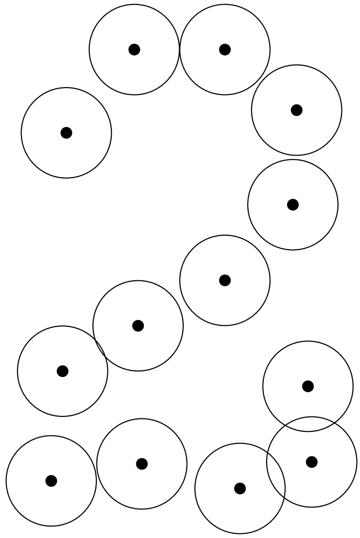

Рис. 2.

звать остовом, он общий для всех изображений класса, и, полагаем для наглядности на данном этапе, присутствует явно в виде точек в каждом изображении, хотя в более общем случае точки остовов можно считать особыми, вспомогательными точками, не обязательно совпадающими с какими-то собственно точками изображений. Весь класс порожденных таким образом изображений можно рассматривать как определяющий преобразования для этих конкретных изображений более широкие, чем аффинные преобразования: каждая точка в пределах круга может быть преобразована в любую другую точку, но в пределах того же круга. Если в остове m точек, то его можно рассматривать как приближение m-го уровня для изображений класса (соответственно называть m-остовом). Остов есть «выжимка» из каждого такого изображения, и эта выжимка у них одинаковая (точнее: остовы аффинно эквивалентны, в таком представлении, как на рис. 2, просто совпадают). Полагаем, что описанные изображения класса можно, с некоторой степенью убедительности, считать относящимися к одному образу. Сразу отметим, что если в исходном изображении n точек, то можно потенциально рассматривать классы изображений, порождаемых уровнями, начиная от 1 и до n, причем изображения n-го уровня будут уже просто попарно аффинно эквивалентными.

Описанное было первым приближением. В следующем приближении можно расширить класс изображений m-го уровня, считая, что в любой из кругов базового изображения (рис. 2) может помещаться любое множество точек, не обязательно с центром именно в центре круга, но, конечно, внутри круга и без пересечений с другими кусками того же изображения. Если исходный рис. 2 рассматривать как футляр, то такие новые изображения можно трактовать как наполнение футляра, и точки наполнения, по сути, ограничены только тем условием, что они находятся внутри кругов. Полагаем, однако, что если есть набор кусков A с m-остовом a (базовое изображение, рис. 2), и набор кусков B с m-остовом b, то a и b находятся в искомом взаиморасположении, то есть точки изображений a и b максимально «придвинуты» друг к другу (в предыдущем, более частном случае, эти остовы просто совпадали). Этим исключается, например, такой случай, когда точки из a и b не совпадают, но только потому, что сдвинуты по отношению друг к другу параллельным переносом.

Итак, куски изображения B из наполнения у нас уже не обязательно равновелики по размерам (ранее равновеликость имела место по причине вписанности кусков в круги одинакового радиуса на рис. 2). Но для кусков исходного изображения A это условие пока сохраняется. Равновеликость важна, ибо в противном случае разбиение трудно считать представляющим форму исходного изображения. Отойти от условия равенства по размерам кусков (для A), сохраняя смысл равновеликости, можно так: пусть разбиение на куски с центрами $a_1, ..., a_m$ таково, что существует R такое, что каждый кусок находится внутри круга радиуса R (и с центрами в $a_1, ..., a_m$), и круги эти не пересекаются (это условие далее уточняется). Смысл изменения в том, что теперь неважно, какого размера собственно кусок исходного разбиения (он может быть даже точкой), но кусок находится внутри круга, одинакового по радиусу с другими кругами, и эти круги не пересекаются. Такой набор кругов с центрами в точках остова называем также футляром (порожденным данным изображением). Уточним условия на радиус кругов и возможное их пересечение. Полагаем, что ни один круг не включает центр другого круга в качестве внутренней точки (т.е. максимальный радиус равен минимальному расстоянию между точками остова). Содержательный смысл этого условия (ограничения) очевиден. (Пример: пусть точка a — центр первого круга, и этот круг включает точку b — центр второго круга. Тогда a можно преобразовать в точку b, и затем в любую точку второго круга, в том числе, и за пределами первого круга, что, по смыслу, не допустимо). Круги могут соприкасаться, и даже пересекаться, но к внутренним точкам круга (усеченного) относим только большую часть круга по соответствующую сторону от отрезка — границы пересечения. Точки наполнения футляра — только внутренние точки (возможно усеченных) кругов.

Как совместить привязку футляров к конкретным изображениям в среде с тем, что мы декларируем рассмотрение с точностью до аффинных преобразований.

Мы намереваемся рассматривать изображения при распознавании с точностью до аффинных преобразований, т.е. безотносительно к параллельным переносам, преобразованиям симметрии, вращениям, изменениям в размерах, сжатиям, растяжениям и любым их комбинациям. Это в какой-то мере соответствует тому, что мы рассматриваем именно форму фигур, если говорить о фигурах, а не какие-то не имеющие отношения к форме обстоятельства, связанные с внешними системами координат [1]. Вместе с тем известно, что в целом аффинные преобразования не сохраняют форму фигур: чрезмерными сжатиями или растяжениями можно сделать фигуру неузнаваемой в сравнении с оригиналом. Тем самым, эта чрезмерность должна быть ограничена. Но чем и как, в каких пределах, и чем эти пределы должны определяться?

Содержательные рассмотрения выше были явно привязаны к «месту», т.е. к конкретному изображению, которое мы называли базовым. Конкретными были и размеры кусков, и радиусы кругов. Как это совместить с рассмотрением с точностью до аффинных преобразований?

Положим в нашей среде S есть группа G изображений «двойки», безусловно вложимых аффинными преобразованиями в футляр A на рис. 2, но разные по размерам, положению на плоскости, с локальными изменениями, и пр., в том числе, для примера, может присутствовать и изображение X, аффинно эквивалентное с изображением из центров кругов на рис. 2, однако сжатое к некоторой прямой с таким коэффициентом, что все его точки выстроились практически в одну прямую. Ясно, что узнать в этом множестве точек X остов из рис. 2 невозможно, хотя они и аффинно эквивалентны. Так вот в наших модельных построениях должно быть учтено, с одной стороны, то, что этот X с остовом на рис. 2 аффинно «совпадают», с другой — то, что уж примером формы «двойки» X явно служить не может.

Изображения группы G — это конкретные наборы точек в разных частях среды S. С содержательной точки зрения они в разной степени соответствуют тому, чтобы называться типичной двойкой, эталоном. Мы должны выделить изображение, наиболее достойное того, чтобы считаться эталоном в группе, причем, напомним, безотносительно к размерам, ориентации и пр., к сжатию-растяжению в определенных пределах, причем эти пределы должны возникать «внутри» модели, а не задаваться извне.

Мы будем пробовать на роль эталона группы поочередно все изображения. Возьмем, для начала, изображение A на рис. 2. Положим, к виду рис. 2 оно приводится разбиением исходного множества точек на m равновеликих кусков, где m — число кругов на изображении рис. 2, и центры кусков — точки изображения рис. 2. Изображение из центров кусков обозначим через a^+. Далее для произвольного изображения A' из группы G осуществляем все возможные его разбиения на m кусков (вообще говоря, уже не заботясь о равновеликости), строим (для каждого разбиения) точечное изображение $a^{+'}$ из центров этих кусков, и укладываем аффинными преобразованиями изображение $a^{+'}$ на изображение a^+. Если укладка (искомое взаиморасположение) такова, что при этом и каждый соответствующий кусок оказывается внутри соответствующего круга, то изображение A' называем приемлемым для эталона A (при уровне дробности m эталона). Поочередно проверяем на приемлемость все изображения группы, и поочередно при каждом изображении, рассматриваемым в качестве эталона. Если приемлемости нет в каждом из этих рассмотрений, то уровень m дробности называем не адекватным группе.

Наибольший уровень адекватности группы называем ее дробностью, он и служит основной характеристикой «однородности» группы, близости по форме составляющих ее изображений. Чем ближе по форме изображения, тем, полагаем, больше дробность. Предельный случай — когда все изображения аффинно эквивалентны — даст максимально возможную в этом случае дробность, равную числу точек в каждом изображении. Минимальная дробность, и она есть всегда, для любой группы, равна единице. Разнородность изображений в группе понижает дробность: ясно, что если, например, в группу «двоек» добавить «четверку», то дробность понизится.

Нетрудно видеть, что в этих построениях хоть и используются в целом аффинные преобразования, но они все же ограничены, в известной степени, примерами, т.е. изображениями группы G. И это разумно, ибо возможных аффинно преобразованных изображений континуум (их можно назвать своеобразными «фантазиями»), но их приемлемость мы связываем с конечным «опорным» множеством, т.е. множеством конкретных примеров из данной среды S, с изображениями группы G (это есть представленная нам «реальность», в отличие от фантазий).

Перейдем к относительно более формальным построениям. Рассмотрим построение на основе произвольного изображения A того, что назовем футляром. Зададимся некоторым m из промежутка от 1 до n (n — число точек в A). Рассмотрим покрытие A' изображения A непересекающимися кусками $A_1, ..., A_m$ (таких разбиений — конечное множество). Центры кусков обозначим через $a_1, ..., a_m$, в целом они составляют изображение a^+, его называем остовом (m-остовом). Пусть R — наименьшее расстояние между точками $a_1, ..., a_m$, полагаем каждую из этих точек центром круга радиуса R, и каждую из точек куска A_i ближе к a_i, чем к другим центрам. Последнее значит, что если круги пересекаются (по отрезку прямой), то кусок должен быть внутри усеченного круга. Если R меньше половины размера покрытия, то разбиение A' называем неправильным. Далее, когда имеются ввиду футляры, рассматриваем только правильные разбиения для A.

Футляр есть пара $<$изображение A, остов $O>$. Этой пары достаточно, чтобы по ней на изображении A построить систему кругов радиуса R с центрами в точках из O, определить границы-отрезки пересечения кругов (если пересечения есть), куски изображения A в каждом из кругов.

Пример (вырожденный): изображение A состоит из точек $a_1, ..., a_n$, рассматриваем разбиение на n кусков, т.е. каждый кусок — точка, R равен минимальному расстоянию между точками.

Итак, построен m-футляр изображения A. Если такой футляр не единственный, то каждый рассматривается независимо.

Пусть теперь B' есть одно из возможных разбиений изображения B на m непересекающихся кусков (не обязательно правильное), с центрами $b_1, ..., b_m$ (обозначение в целом: b^+). Если мы теперь аффинными преобразованиями трансформируем изображение B' вместе с точками $b_1, ..., b_m$, то новые положения точек $b_1, ..., b_m$ будут по прежнему центрами преобразованных кусков. Это следует из теоремы 1.

Теорема 1. *Пусть дано изображение X из точек $x_1, ..., x_n$ и точка y — его центр. Пусть изображение X' из точек $x'_1, ..., x'_n$ есть аффинно преобразованное изображение X, и точка y' — его центр. Тогда изображения из точек $x_1, ..., x_n, y$ и точек $x'_1, ..., x'_n, y'$ аффинно эквивалентны.*

Доказательство. Утверждение означает, что при аффинных преобразованиях изображения X из точек $x_1, ..., x_n, y$ в изображение X' из точек $x'_1, ..., x'_n, y'$ центр y изображения X переводится в центр y' изображения X'. Действительно, допустим, что центр преобразованного изображения X' есть точка y'', отличная от y'. Точка y есть центр эллипса E минимального по площади, в который вписано изображение X, т.е. если обозначить площадь эллипса через $P(E)$, площадь выпуклой оболочки изображения X через $P(X)$, то отношение $P(E)/P(X)$ минимальное из возможных. Такой эллипс согласно [5] существует и единственен. Аналогично точка y'' есть центр наименьшего по площади эллипса E'' для изображения X'. При аффинном преобразовании эллипс E преобразуется в эллипс E' с центром в y'. Поскольку y' и y'' предполагаются разными, то и эллипсы E' и E'' тоже разные. Но при аффинном преобразовании отношение $P(E)/P(X)$ сохранится равным $P(E')/P(X')$, а это значит, что эллипс E' тоже минимален по площади для X' и единственен — пришли к противоречию. Итак, E' и E'' должны совпадать, а, значит, совпадают и y'' с y'. Теорема доказана.

□

Определим понятие вместимости для B' в m-футляр изображения A. Расположим изображение b^+ на изображении O искомым образом. Это преобразует и изображение B' в целом. Если при этом каждый кусок изображения B' окажется внутри соответствующего круга (возможно, усеченного) футляра изображения A, то говорим, что данное m-разбиение изображения B вместимо в этот футляр. Изображение B называем m-вместимым в изображение A, если хотя бы одно из его m-разбиений вместимо в хотя бы один из m-футляров изображения A.

Понятие дробности для группы. Пусть группа G из $S^\star$ состоит из изображений $A_1, ..., A_k$, и пусть u — наименьшее число точек в изображениях группы. Зададимся некоторым m из промежутка от 1 до u и пусть существует хотя бы одно из изображений группы такое, что все $A_1, ..., A_k$ m-вместимы в некоторый футляр этого изображения. Тогда группу называем m-совместимой. Наибольшее значение m, для которого группа m-совместима, обозначаем через M и называем дробностью группы. Те из изображений группы, M-футляры которых вмещают все изображения группы, называем ее эталонами. Теперь каждая группа из $S^\star$ снабжена характеризующим ее параметром — дробностью, набором эталонов, и, будем полагать, M-футлярами на эталонах.

Пусть G_1 и G_2 — группы из $S^\star$ с одинаковой дробностью, и пусть множество изображений G_2 есть подмножество (собственное) для группы G_1. Тогда говорим, что группа G_1 полнее группы G_2. Группу называем полной, если нет группы полнее, чем она. Таким образом, здесь понятие полной группы основывается на задании группы перечислением входящих в нее изображений.

Пусть G — группа, M' есть M-футляр одного из эталонов группы. По построению, все изображения группы вместимы в M'. Однако, в M' могут быть вместимы и другие изображения среды. Назовем совокупность всех вместимых в M' изображений максимальным наполнением футляра M'.

Теорема 2. *Если G — полная группа, то максимальное наполнение любого из M-футляров группы совпадает с группой. Однако если максимальное наполнение M-футляров группы совпадает с группой, то из этого не следует, что группа полная.*

Доказательство. Первое утверждение состоит в том, что если G — полная группа, то максимальное наполнение любого из футляров группы совпадает с группой. Действительно, пусть это не так, и существует изображение X, которое не входит в G, но принадлежит максимальному наполнению одного из футляров M'. Но тогда расширенная группа G с присоединенным изображением X будет M-совместимой группой, что противоречит исходной полноте группы G.

Вторая часть теоремы утверждает, что обратное неверно, т.е. если максимальное наполнение футляров группы совпадает с группой, то из этого не следует, что группа полная. Докажем это построением примера. Рассмотрим класс g некоторых аффинно эквивалентных изображений из $S^\star$, причем всех таких изображений. Пусть x — одно из этих изображений, причем на основе этого изображения строится футляр группы. Полагаем, в x содержится m точек (по меньшей мере четыре), все они — точки остова, и две из них, a_1 и a_2, наиболее близко расположенные, определяют величину R футляра, равную некоторой величине эпсилон. Если эпсилон достаточно близок к нулю, то в этот футляр вместятся изображения группы g и только они, то есть максимальное наполнение футляра совпадает с группой. Построим на основе изображения x изображение x', совпадающее с x, за исключением точек a_1 и a_2, замененных на точки соответственно a'_1 и a'_2, расположенные на продолжении отрезка $(a_1 a_2)$ и симметрично относительно середины этого отрезка. Все точки изображения x' считаем точками футляра, точки a'_1 и a'_2 — наиболее близкими друг к другу, и величину R футляра полагаем равной четырем эпсилон. При достаточно малом эпсилон в футляр на изображении x' будут вмещаться все изображения группы g и только они. Однако изображение x' в футляр на x вместиться, по построению, не может. Группа G, состоящая из изображений группы g и присоединенного изображения x', тем самым, m-совместима, значит группа g не полная, что и доказывает утверждение.

$\square$

Итак, ранее для полной группы было возможно ее задание, как и для всех групп, перечислением всех входящих в группу изображений. Не очень удобный способ, если изображений много. Однако теперь для полной группы возможно ее задание каким-либо футляром эталона группы, и указанием, что группу составляют все вместимые в этот футляр изображения.

Множество всех полных групп обозначаем через $S^{\star\star}$ и называем множеством образов первого рода. Возможность задавать образ первого рода через посредство футляра эталона можно считать способом компактного описания образа. Итак, перейдя от $S^\star$ к $S^{\star\star}$ мы сократили («ужали») множество групп, потенциально рассматриваемых как образы, причем, скорее всего, существенно. Однако доказательно оценить разницу по мощности множеств $S^\star$ и $S^{\star\star}$ будет, возможно, нетривиальной задачей.

Как можно содержательно трактовать описанное выше, т.е. наличие полной группы G и ее подгрупп, неполных, но максимальное наполнение которых совпадает с G? Задание примеров изображений группами (множество $S^{\star}$) означает, что эти подгруппы могут быть разными: где-то примеров больше, где-то меньше, причем для одного и того же образа. В среде, в разных ее частях, может быть много разных примеров одной и той же фигуры (возможно, с вариациями в контурах). Скажем, некто познакомился с изображением «двойки» на сравнительно небольшом множестве примеров (т.е. это одна из возможных групп в $S^{\star}$). Он «выработал» футляр для этой группы и это позволит ему при появлении неизвестного изображения определить, вместимо оно в этот футляр или нет, то есть распознать изображение. Другой субъект «выработает» схожий футляр (иными словами, схожее понятие «двойки») на другом множестве ее примеров (другая группа), но в целом, возможно, изображения первой группы вместимы в футляр второй и наоборот, что и означает, что их можно объединить в одну группу. В целом это представляет в модели то содержательно очевидное обстоятельство, что одному и тому же образу можно «обучиться» на разных примерах и из разных частей среды.

О распознавании применительно к образам первого рода. Пусть задано в среде S некоторое изображение X, трактуемое как неизвестное, и его нужно отнести к образам в $S^{\star\star}$, т.е. выяснить, в каких полных группах это изображение содержится. Это, по сути, аналог (несколько расширенный) традиционной постановки задачи распознавания.

Представим образы (группы из $S^{\star\star}$) вершинами ярусного графа с номерами ярусов от 1 до N. Это значит, что каждой вершине приписана группа как таковая, а также футляры эталонов групп (т.е. группа задана обоими способами: и перечислением входящих в нее изображений, и футляром). На ярусе с номером i располагаются все вершины с дробностью i. Из вершины X яруса i идет ребро в вершину Y яруса j, где $i < j$, если группы X и Y имеют непустое пересечение, и нет группы Z из яруса с номером k, где $i < k < j$, такой, что Z имеет непустые пересечения и с X, и с Y. Вершины X и Y называем смежными, причем X — предковая вершина, Y — потомковая.

Начинаем проверку принадлежности распознаваемого изображения X образам с яруса 1. На этом ярусе образ только один, и любой X ему всегда принадлежит. Далее действуем в рамках индуктивной процедуры: если выяснили, что X принадлежит вершинам (образам) $X_1, ..., X_k$ на ярусе с номером i, то в качестве следующего шага надо проверить на принадлежность к ним только те образы на ярусах с большими номерами, которые являются потомковыми по отношению к вершинам $X_1, ..., X_k$. Каждая вершина на ярусном графе задана двояко, во-первых, как группа изображений, во-вторых, футляром какого-либо эталона группы. Для проверки принадлежности X образу надо либо выяснить принадлежность X группе, либо выяснить вместимость X в группу с использованием футляра. Принадлежность группе означает проверку на аффинное совпадение со всеми изображениями группы, т.е. перебор. Если в группе большое число (скажем, многие миллионы) изображений, то это процесс затратный и небыстрый. Рациональнее определять вместимость. При этом, однако, при проверке на вместимость X в некоторый футляр, надо определять искомое взаиморасположение остова, полученного при разбиении изображения X на куски, с остовом футляра образа, а это, в общем случае, перебор $n!$ вариантов биекций остовов (что тоже немало), где n — число точек в каждом из них.

Сократим перебор. Здесь мы можем опереться на теорему 3, которая будет доказана ниже. Содержательный смысл состоит в том, чтобы, при движении по ярусам графа, распознавание, уже осуществленное для образа из яруса с меньшим номером, использовать для уменьшения перебора при выяснении принадлежности изображения образу на ярусе с большим номером.

Пусть для группы A яруса i существует футляр $F^i(A)$ с остовом O^i (точки $o_1, ..., o_i$), и пусть для изображения a из A его разбиение $a_1, ..., a_i$ — из числа вместимых в футляр $F^i(A)$, причем с соответствием куска a_k точке $o_{\varphi_1(k)}$ $(k = 1, ..., i)$. Тогда точки изображения a из куска a_k считаем соответствующими точке $o_{\varphi_1(k)}$ остова O^i.

Пусть для группы B яруса j $(i < j)$, смежной с группой A, существует футляр $F^j(B)$ с остовом O^j (точки $o'_1, ..., o'_j$), и пусть для изображения a из B его разбиение $a_1, ..., a_j$ — из числа вместимых в футляр $F^j(B)$, причем с соответствием куска a_t точке $o'_{\varphi_2(t)}$ $(t = 1, .., j)$. Тогда точки изображения a из куска a_t считаем соответствующими точке $o'_{\varphi_2(t)}$ остова O^j.

Точку $o_{\varphi_1(k)}$ остова O^i называем соответствующей точке $o'_{\varphi_2(t)}$ остова O^j (и наоборот), если a_k и a_t имеют непустое пересечение. Этим определено соответствие ψ между точками остовов O^i и O^j. Оно, естественно, не взаимно однозначное. Рассматриваем все такие соответствия, порождаемые всеми изображениями, общими для A и B и всеми их разбиениями, в их искомых взаиморасположениях с остовами всех футляров (эталонных) для A и B, множество этих соответствий обозначаем через $\psi(O^i, O^j)$. Множество $\psi(O^i, O^j)$ считаем нагрузкой ребра между смежными вершинами A и B.

Содержательный смысл множества $\psi(O^i, O^j)$ состоит в следующем. Вершины A и B на графе смежные. Значит, группы A и B имеют общие элементы, или, что практически то же самое, есть изображения, вместимые одновременно и в A, и в B. Пусть x — такое изображение. Пусть x с разбиением $x_1, ..., x_i$ вмещается в футляр $F^i(A)$ с остовом O^i из точек $o_1, ..., o_i$ с соответствием куска x_k точке $o_{\varphi_1(k)}$ $(k = 1, ..., i)$. Пусть теперь x с разбиением $x'_1, ..., x'_j$ вмещается в футляр $F^j(B)$ с остовом O^j из точек $o'_1, ..., o'_j$ с соответствием куска x'_t точке $o'_{\varphi_2(t)}$ $(t = 1, ..., j)$. Задаем соответствие между разбиениями $x_1, ..., x_i$ и $x'_1, ..., x'_j$ условием: x_k соответствует x'_t (и наоборот), если x_k и x'_t имеют непустое пересечение. Это, очевидно, дает и соответствие ψ между точками остовов O^i и O^j. Соответствие ψ по построению принадлежит $\psi\left(O^i, O^j\right)$. Это, однако, не значит, что если, например, мы возьмем другие вместимые разбиения $x_1, ..., x_i$ и $x'_1, ..., x'_j$, то ψ будет другим. При разных разбиениях и разных футлярах групп порождаемые ими соответствия могут совпадать или не совпадать. Но $\psi\left(O^i, O^j\right)$ — это множество всех возможных соответствий, которые порождаются (реализуются) в рамках описанного. Можно предположить, что число так реализуемых соответствий в общем случае не слишком большое (в сравнении с числом всех возможных соответствий).

Пусть теперь Z — изображение, вместимое в A, и не обязательно вместимое в B. Это значит, Z вместимо в группу A, т.е. что для группы A яруса i, для футляра $F^i(A)$ с остовом O^i (точки $o_1, ..., o_i$), и для изображения Z его разбиение $z_1, ..., z_i$ — из числа вместимых в футляр $F^i(A)$, причем с соответствием куска z_k точке $o_{\varphi_1(k)}$ $(k = 1, ..., i)$. Тогда точки изображения Z из куска z_k считаем соответствующими точке $o_{\varphi_1(k)}$ остова O^i.

Пусть для изображения Z из B его разбиение $z'_1, ..., z'_j$ соотнесено с футляром $F^j(B)$, причем с соответствием куска z'_t точке $o'_{\varphi_2(t)}$ $(t = 1, ..., j)$. Подчеркнем, что нам не известно, вместимо ли Z в B при соответствии $\varphi_2(t)$. Пусть теперь некоторое соответствие ψ' между точками остовов футляров O^i и O^j задано тем условием, что z_k и z'_t имеют непустое пересечение: в этом случае точку $o_{\varphi_1(k)}$ полагаем соответствующей точке $o'_{\varphi_2(t)}$ (и наоборот). Если ψ' принадлежит $\psi\left(O^i, O^j\right)$, то $\varphi_2(t)$ называем потенциально приемлемым соответствием между точками остова O^j и кусками разбиения изображения Z. Обозначаем через $\{\varphi_2(t)\}$ множество всех потенциально приемлемых соответствий, полученных при всех имеющихся разбиениях $z_1, ..., z_i$, соответствиях $\varphi_1(k)$, всех разбиениях $z'_1, ..., z'_j$, и всех $\varphi_2(t)$.

Теорема 3. *Если Z вместимо в B, то существует его разбиение $z'_1, ..., z'_j$ с соответствием $\varphi_2(t)$ кусков этого разбиения точкам остова O^j футляра образа B такое, что соответствие $\varphi_2(t)$ принадлежит $\{\varphi_2(t)\}$.*

Доказательство. Доказательство следует из представленных выше построений, из которых явствует, что соответствие $\varphi_2(t)$, при котором изображение Z вместимо в B, является с необходимостью потенциально приемлемым.

Таким образом, при выяснении вопроса о вместимости изображения Z в B нет нужды перебирать все возможные биективные соответствия между кусками разбиения $z'_1, ..., z'_j$ и точками остова O^j (их потенциально $(j!)$), достаточно ограничиться множеством $\{\varphi_2(t)\}$ потенциально приемлемых соответствий.

Отметим, что имеется упрощенная демонстрационная реализация алгоритма распознавания в виде компьютерной программы.

$\square$

Список литературы

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

[2] Козлов В. Н., "О зрительном образе, математических подходах к определению этого понятия и о распознавании изображений", *Журнал вычислительной математики и математической физики*, **39**:11 (1999), 1929–1946.

[3] Kozlov V. N., "Visual Pattern and Geometric Transformation of Images", *Pattern Recognition and Image Analysis*, **10**:3 (2000), 321–342.

[4] Данцер Л., Грюнбаум Б., Кли В., *Теорема Хелли и ее применения*, М.: Издательство «Мир», 1968.

[5] Загускин В. Л., "Об описанных и вписанных эллипсоидах экстремального объема", *Успехи математических наук*, **13**:6 (1958), 89–93.

On the definition of the concept of a visual image

Kozlov V.N.

The work belongs to the field of theoretical aspects of visual image recognition. The images are treated as finite sets of points in Euclidean spaces of various dimensions. This is how real images can actually be represented. There is no formal definition of a visual image yet. The article makes a first approximation to this definition.

Keywords: image recognition, visual images.

References

[1] Kozlov V. N., *Vvedenie v matematicheskuyu teoriyu zritel'nogo vospriyatiya [Introduction to the mathematical theory of visual*

perception], Publishing House of the Center for Applied Research at the Faculty of Mechanics and Mathematics of Moscow State University, Moscow, 2007 (in Russian).

[2] Kozlov V. N., "O zritel'nom obraze, matematicheskikh podkhodakh k opredeleniyu ehtogo ponyatiya i o raspoznavanii izobrazhenii [On the visual image, mathematical approaches to the members of this concept and on image recognition]", *Computational Mathematics and Mathematical Physics*, **39**:11 (1999), 1929–1946 (in Russian).

[3] Kozlov V. N., "Visual Pattern and Geometric Transformation of Images", *Pattern Recognition and Image Analysis*, **10**:3 (2000), 321–342.

[4] Danzer L., Grunbaum B., Klee V., *Helly's theorem and its relatives*, Amer. Math. Soc. Providence, RI, 1963.

[5] Zaguskin V. L., "Ob opisannykh i vpisannykh ehllipsoidakh ehkstremal'nogo ob"ema [On circumscribed and inscribed ellipsoids of extremal volume]", *Russian Mathematical Surveys*, **13**:6 (1958), 89–93 (in Russian).