

УДК 519.7

В.А. Утробин

**ЭЛЕМЕНТЫ ТЕОРИИ АКТИВНОГО ВОСПРИЯТИЯ ИЗОБРАЖЕНИЙ**

Нижегородский государственный технический университет им. Р.Е. Алексеева

Излагаются основные положения теории активного восприятия как информационной модели врожденных механизмов системы зрительного восприятия этапа предварительной обработки изображений в условиях априорной неопределенности.

*Ключевые слова:* зрительное восприятие, распознавание образов, априорная неопределенность, анализ изображений, информационные модели, этапы обработки изображений.

**Введение**

Более девяноста процентов информации об окружающем мире человек получает через систему зрительного восприятия. Эта информация необходима для жизнеобеспечения человека и ориентации в пространстве. Поэтому проблема зрительного восприятия является одной из фундаментальных проблем современности. Возникнув на заре прогресса человеческой мысли, она остается актуальной и в настоящее время. Тем не менее, несмотря на глубокие исторические корни, изображение стало предметом точных наук лишь в середине пятидесятых годов прошлого столетия и причиной тому явилось бурное внедрение методов кибернетики в задачи моделирования биосистем. Была высказана гипотеза – механизм восприятия есть классифицирующая система, и сформулирована задача построения машины, способной обучаться. Результатом такой общей постановки проблемы являются два крупных взаимосвязанных направления исследований – разработка математических моделей зрительного восприятия и разработка математических методов информационных преобразований изображения как «простого» многомерного сигнала, породившая современную теорию распознавания образов, заключающуюся в построении математических моделей классификации объектов в режиме обучения и не учитывающую специфику изображения как объекта исследования.

Основная задача распознавания (расознавания в «узком» смысле) заключается в построении правила, позволяющего отнести наблюдаемый объект, заданный на системе признаков (параметров), к одному из классов, удовлетворяющих требованиям сходства – максимум однородности внутриклассовых объектов при максимуме различия на межклассовых объектах. Пусть  $\mathbf{x} = \{x_i; i = 1, \dots, d\}$  – вектор наблюдаемого объекта в евклидовом пространстве  $E^d$ , называемом пространством признаков (объектов);  $T = \{T_i; i = 1, \dots, d\}$  – множество признаков, позволяющих получить решение о наличии объекта;  $R = \{R_i; i = 1, \dots, R\}$  – множество классов, удовлетворяющих требованиям сходства, каждому элементу которого принадлежит множество точек из  $E^d$ ;  $G = \{g_i(\mathbf{x}); i = 1, \dots, R\}$  – множество правил (функций), отображающих наблюдаемый объект  $\mathbf{x}$  на  $R$ . Тогда система распознавания образов есть тройка  $S = (T, R, G)$ , а процесс распознавания, есть отображение на классе функций вида  $G: \mathbf{x} \rightarrow R$ . В этом случае на многообразии известных систем распознавания можно выделить следующие классы [1]:

- *в зависимости от способа выбора правила* – распознавание с обучением (просто распознавание), либо самообучение (автоматическая классификация);
- *в зависимости от делимости классов* – детерминистский, либо вероятностный подход;

- в зависимости от качественного состава множества  $T$  – дискриминантный, либо структурный подход;
- в зависимости от априорного знания о множествах  $T$ ,  $R$  – распознавание в узком смысле, либо распознавание в широком смысле, как распознавание в условиях априорной неопределенности.

Именно для последнего класса систем до настоящего времени отсутствовала необходимая теория распознавания.

*Введем ряд определений.* 1. Система, реализующая задачу распознавания в условиях априорной неопределенности, есть **универсальная система распознавания**. 2. Пусть каким-либо образом априори определены: все возможное (конечное) множество объектов, конечное множество признаков; критерий однородности внутри класса (кластера, таксона, образа). Тогда универсальная система распознавания – **самообучающаяся система распознавания**.

Решая задачу декомпозиции проблемы распознавания с позиций системного анализа, можно выделить три основных этапа преобразований [1]: – этап формирования исходного описания (с позиций системного анализа – этап подготовки данных), этап формирования признаков (этап анализа) и этап классификации (этап принятия решения). Если последний этап информационных преобразований теоретически «хорошо проработан», то первые два этапа в системах, базирующихся на теории распознавания образов, «игнорируются». Тем не менее, в системе зрительного восприятия эти два этапа информационных преобразований занимают важное место и врожденны, т.е. «не отягощены» действием сознания, а этап принятия решения входит в состав этапа понимания. При этом зрительная система «функционирует» как самообучающаяся система в условиях априорной неопределенности, обеспечивая восприятие увиденного в любой момент времени при любых неизвестных условиях, включая организованные помехи и загромождения.

Под *восприятием* вообще и зрительным восприятием в частности понимают процесс чувственно-исследовательской деятельности, направленный на ознакомление с предметом или явлением, воздействующим на органы чувств, конечным результатом которого является отражение исследуемого объекта в некоторую совокупность его свойств, описывающих его объективную целостность. Восприятие как информационный процесс наделяется следующими объективными свойствами:

1. Целостность – изображение (как объект исследования) есть целостное образование. Тогда задача этапа анализа состоит в декомпозиции целого на части с образованием уровней представлений изображения.

2. Структурированность – стратификация процесса восприятия на страты: обнаружения, локализации, опознания, интерпретации.

3. Стратегия (переключение внимания) восприятия – распространение процесса восприятия от более важных элементов изображения к менее важным.

4. Обобщенность оснований – реализация процесса восприятия, обеспечивающая выделение максимально устойчивых признаков объектов и их отношений на изображении в условиях априорной неопределенности последнего.

5. Стереотипность – обеспечение процесса восприятия однотипностью (однородностью по процедуре формирования) и однородностью (по составу) элементов и связей.

Выделенным свойствам системы зрительного восприятия как самообучающейся системы соответствуют информационные процессы преобразования зрительной информации на основе теории активного восприятия.

### Модели объекта исследования

Любая система восприятия, по определению, решает задачу отражения предмета – объекта окружающего мира, в совокупность его свойств с учетом его объективной целостности. Результатом отражения является (перцептивный) образ, а сам процесс отражения – результат взаимодействия замкнутой системы: окружающий мир и наблюдатель, погруженный в этот

мир и неискажающий его. Под *окружающим миром* будем понимать пространство входных (для наблюдателя) величин (воздействий), принадлежащих трехмерному евклидову пространству на множестве моментов времени. Под *наблюдателем* будем понимать систему, состоящую из двух подсистем – воспринимающей (регистрирующей) и интерпретирующей (обрабатывающей). *Первая* решает задачу отображения любой доступной наблюдателю функции воздействия в функцию выхода  $\mu(x, y, z, t)$ : если входное воздействие в момент времени  $t$  есть функция интенсивности светового потока в видимой части спектра, то  $\mu(x, y, z)$  в тот же момент времени – психофизиологическая *функция яркости*; именно эта функция, принадлежащая некоторому множеству  $M$ , есть изображение. *Вторая* подсистема реализует отображение любого изображения в *функцию выхода*, которую назовем *описанием произвольного изображения*, принадлежащего  $M$ . Такое представление позволяет выделить ряд (аксиоматических) свойств взаимодействующих подсистем в составе замкнутой системы. Свойства окружающего мира, представленного  $E(x, y, z, t)$ : *наблюдаемость*, *измеримость*, *открытость*, *упорядоченность* и *структурированность* (раскрытие перечисленных свойств можно найти в работах [1, 2]). Свойства подсистем наблюдателя: 1) *эквивалентность*, т.е. воспринимающая подсистема не должна искажать функцию воздействия; 2) *ограниченность*, т.е. для воспринимающей подсистемы: а) объем информации об окружающем мире ограничен наблюдаемым пространством (полем зрения) и допустимым динамическим диапазоном; б) воспринимающая подсистема есть система проектирования трехмерной функции воздействия на двумерную поверхность рецепторов; в) функция выхода  $\mu(x, y)$  априори дискретна в пространстве, так как дискретен характер взаимодействия окружающей среды с конечным дискретным множеством рецепторов. Тем не менее, разрешающая способность, например, зрительного анализатора (сетчатки) так велика, что  $\mu(x, y)$  допустимо рассматривать как непрерывную функцию координат. Для обрабатывающей подсистемы как любой физической системы свойство ограниченности замкнуто на свойство конечности пропускной способности.

Следовательно, под изображением будем понимать множество, каждый элемент которого в фиксированный момент времени  $t$  есть неотрицательная действительная функция действительных аргументов вида  $M_t = \mu(x, y)$ , определенная на конечном множестве точек замкнутой двумерной области определения (поле зрения), погруженной в евклидово пространство, суммируемая, квадратично интегрируемая на этом множестве и наделенная выше перечисленными свойствами. Так определенный объект исследования позволяет выделить два понятия: пространственный, вневременной объект (статическое, «замороженное» изображение, просто изображение) – это функция  $M_t = \mu(x, y)$ , определенная в момент  $t$ ; пространственно-временной объект (динамическое изображение) – это функция  $M_d = \varphi(t, M_t)$ . Поскольку изображение  $M_t \in M$  допускает дискретизацию, то оно допускает представление на решетчатую (сеточную) функцию на прямоугольной области определения  $L \times N$  (в условиях априорной неопределенности такая область допустимо квадратная область –  $N \times N$ ). Отсюда следует, что мощность множества изображений конечна и равна  $k_{\max} = P^{N \times N}$ , где  $P$  – число градаций яркости.

Таким образом, все многообразие изображений существующих в природе хоть и велико, но конечно, а значит, система, раскрывающая неопределенность наблюдаемого изображения, есть автомат без памяти – «простая» *комбинационная схема*. Тем самым доказано первое условие на определение самообучающейся системы. Более того, представление изображения в виде пространственной модели  $\mu(x, y)$  с его аксиоматическими свойствами по определению позволяет построить изоморфные модели объекта исследования с позиций алгебры моделей:

1) множество  $M$  образует группу  $G = \langle M; + \rangle$  и является подпространством  $E(M)$  скалярного поля  $E$ , где  $E(M)$  – тело действительных чисел на бинарных операциях  $m: M \times M \rightarrow M$ ,  $m = \{+, \times\}$ ;

2)  $M$  есть подпространство  $B(M)$  векторного пространства  $B$  над телом  $E(M)$  с метрикой  $d(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|$ ;

3)  $M$  есть подпространство  $\Gamma(M)$  топологического пространства  $\Gamma$  с базой на семействе  $d(a, b) < \delta$ , изоморфное на классы упорядоченных векторных пространств и направленных векторных решеток;

4)  $M$  есть структура  $L = \langle M; \cup, \cap, e_1, e_0 \rangle$ , если на  $M$  определены бинарные операции  $\cup, \cap$  с соответствующей аксиоматикой, и отношение порядка  $\rho (\leq)$ .

### Модель неопределенности

Пусть каждой точке из  $M = \{\mu(i, j)\}$ , где  $(i, j) \in X \times Y$ , соответствует вероятность элементарного события так, что выполняются все необходимые аксиомы. Поскольку конечное множество  $M$  есть полное пространство элементарных событий – генеральная совокупность в конечном объеме, то если каждому значению  $\mu(i, j) \in M$  соответствует вероятность  $P(\mu(i, j))$ , то  $M$  есть *достоверное* событие на множестве равновероятных элементарных событий  $P(\mu(i, j)) = p \forall \mu(i, j) \in M$ ; отсюда следует, что в условиях априорной неопределенности каждый элемент изображения равновероятен, а отношение между любой парой элементов в области определения есть отношение эквивалентности  $(\mu(i, j), \mu(k, l)) \in \varepsilon$ .

По У. Эшби сложность системы произвольной природы (без учета структурных связей) может быть оценена ее разнообразием – энтропией *неслучайного события* (по А. Н. Колмогорову), состоящего из  $k$  элементов. Пусть  $k$  есть максимально возможное число событий в пространстве изображений. Тогда количество информации, содержащееся в любом  $M$ , равно  $J = \log k$ , которое с позиций информационно-теоретического подхода интерпретируется как априорная неопределенность состояния  $M$ , т.е. максимальная энтропия равна  $H_m$ , либо как количество информации, необходимой для разрешения этой неопределенности, т.е.  $(H_m - H) = O_H$  – организация системы по К. Шеннону, где  $H$  – текущая энтропия.

Обобщением изложенного является модель неопределенности как четырехфакторная модель, учитывающая (в отличие от известных) вероятностную, энтропийную, организационную (в смысле структурной сложности) и пространственно-структурную упорядоченность:

$$N_h = \langle \langle p_i = p \forall i ; H = H_m = \log k ; O_h = \emptyset ; (a_i, a_j) \in \varepsilon \forall i, j \rangle \rangle. \quad (1)$$

Её обратной моделью является модель полной определенности. При этом, если  $N_h$  описывает полностью неопределенную систему – «черный ящик», то обратная модель  $N_o$  – описание полностью известной системы – «белого ящика». Процесс перехода от  $N_h$  к  $N_o$  за один шаг преобразования – *оптимальный процесс раскрытия априорной неопределенности*.

### Модель процесса раскрытия неопределенности наблюдаемого изображения

Поскольку изображение как множество положительно определено и конечно, то оно само и каждая его подобласть допускает интегральное отображение типа ортогонального проектирования (теорема необходимости [1, 2]):

$$m(G_i) = \iint_{G_i} \mu(x, y) dx dy \quad (2)$$

по любой подобласти  $G_i \subseteq G$ .

С позиций физики (2) имеет смысл *массы*, т.е. изображению в области его определения и любой его подобласти допустимо поставить в соответствие массу как меру; эта масса есть «визуальная» масса изображения. Процесс раскрытия неопределенности, в силу (2), реализуется за один шаг преобразования, поэтому процесс оптимален по (1), а так как для его реализации требуется только операция суммирования (в дискретном варианте), то он – *максимально возможно вычислительно прост*. Поскольку весовая матрица преобразования имеет вид  $1(x, y)$  по любой  $G_i$ , то такое преобразование обеспечивает: а) отсутствие необходимости ап-

риорного знания весовой маски фильтра; б) отсутствие необходимости реализации операции свертки, как самой вычислительно сложной операции во всех стандартных методах обработки изображений; в) результат преобразования принадлежит пространству действительных чисел, а не мнимых, как в преобразованиях Фурье.

Преобразование (2) определено **Q-преобразованием**, оно реализует информационные преобразования этапа *исходного описания*, а его интерпретации замыкается на интерпретации отображения пространства на точку массы  $m(G_i)$ :

1) с позиций системного анализа – структурная точка, достоверно (с вероятностью 1) представляющая область  $G_i$  с функцией  $\mu(x, y)$  в этой области;

2) с позиций формальной теории композиции – композиционный (информационно значимый) центр; в) с позиций «точностного» анализа –  $\varepsilon$ -точная (по А.Н. Колмогорову) оценка многомерной функции в области ее определения, обеспечивающая минимум средней квадратической погрешности;

3) с позиций теории управления – ультрастабильная (по У. Эшби) точка и центр «потенциальной ямы»;

4) с позиций математической физики (теории сплошной среды, теории гармонических функций, теорема Гаусса) – средняя функция, допускающая два модельных (восстанавливающих) представления: а) с учетом гармоничности в замкнутой области определения – равномерное распределение потенциала любой точки  $A$  в области с центром  $A_0$ :  $\varphi(A) = m(A_0)$ ; б) с учетом гармоничности в открытой (двумерной) области определения – логарифмическое распределение потенциала относительно тяготеющей массы (или заряда) в центре:  $\varphi(A) = m(A_0) \ln r$ .

Поэтому Q-преобразование реализует отображение изображения, как объекта исследования, в абсолютно (бесконечно) гладкое многообразие.

Поскольку на множестве  $M$  существует отношение эквивалентности, то  $M$  факторизуемо, т.е. допускает естественный процесс разбиения на непересекающиеся открытые, либо замкнутые (гладко склеенные топологически) подобласти (карты). Оптимальным процессом разбиения в условиях априорной неопределенности является процесс дихотомии. Поэтому, разделив область  $G$ , где определено изображение, пополам и применив к каждой подобласти процедуру (2), получим две точки с массами  $m(G_1)$ ,  $m(G_2)$ . Отношение между ними как мерами есть отношение  $v(G) = m(G_1) - m(G_2)$ , где преобразование  $v(G) \equiv v(M)$  есть:

1) обратное преобразование относительно (2) (теорема достаточности, [1, 2]);

2) частная производная (а «не просто» конечная разность) по направлению  $x^i$ , ортогональному границе дихотомии области, и компонента градиента

$$v_i(G) = m(G_1) - m(G_2) = \partial\varphi(x, y)/\partial x^i = \mu_i; \quad (3)$$

3) выявление бинарного отношения строгого порядка (либо равенства, симметрии, инвариантности) на паре структурных точек, реализуемое только через арифметическую разность.

Отношение (3) определяет структурную (информационную) связь между структурными элементами системы *изображение*, получаемая за один шаг преобразования и обеспечивающая получение максимума информации в один бит (1). Само отношение применяется к гладкому (абсолютно) многообразию, где «нет помех и где вообще отсутствует понятие – помеха», выявляя дифференциальную структуру на этом многообразии (интегральную кривую).

Пара преобразование, первое из которых – интегральное, второе – дифференциальное, определены **U-преобразованием**, как композицией «истинно» (в смысле последовательности реализации) интегрально-дифференциального преобразования, в отличие от всех известных.

Пара преобразований (2), (3) полна в смысле теоремы Стокса и образует алгебраическую группу, которая, в силу конечности множества  $M$ , порождает пару конечных компози-

ционных рядов прямого и обратного преобразований. Эти ряды изоморфны в силу теоремы Жордана-Гёльдера и реализуемы (теорема реализуемости [1, 2]):

1) преобразование (2) реализуемо на линейном операторе  $w_i = 1(x, y)$  по  $G_i \subseteq G$ ;

2) преобразование (3) реализуемо на линейном операторе  $v_i = \{1(x, y); -1(x, y)\}$ , где первая половина принадлежит одной области дихотомии, вторая – другой.

Порядок рядов равен 16. Такой порядок композиционных рядов с позиций теории булевых алгебр на паре  $(x, y)$  независимых ортогональных направлений пространства  $E^2$ , в соответствии с теоремой А.Н. Колмогорова (о числе булевых функций), свидетельствует о существовании 15 (если исключить из рассмотрения нуль-направление, единицу группы) независимых направлений  $x^i$  дихотомии области определения. Эти же направления, с позиции теории алгебр Ли и полей Киллинга, говорят о существовании 15 направлений сохранения физической величины (свойство сохраняемости обеспечивается инвариантностью скалярной функции относительно некоторого движения и необходимостью, чтобы это движение реализовалось вдоль векторного поля Киллинга того евклидова пространства, в котором реализуется движение; такие поля – поля постоянства фундаментального метрического тензора).

Таким образом, число дихотомий области определения, если не считать первую – нулевую дихотомию – саму область, равно 15; из них:

1) три варианта – дихотомия по  $x$ , дихотомия по  $y$ , дихотомия по  $x$  и  $y$ , определяют и задают матрицу  $2 \times 2$  (пусть это будет матрица уровня  $K_2$ ; тогда преобразованию (2) по всей области соответствует матрица  $1 \times 1$  уровня  $K_1$ ; пусть этой матрице соответствует фильтр  $F_0$ )

и три фильтра  $F_1, F_2, F_3$  с преобразованиями (в  $E^2$ )  $\partial\varphi(x, y)/\partial x$ ,  $\partial\varphi(x, y)/\partial y$ ,  $\partial^2\varphi(x, y)/\partial x\partial y$ ;

2) две пары фильтров по  $x$  и  $y$  соответственно реализуют преобразования по этим направлениям с наращиванием порядка производной –  $\{\partial^2\varphi(x, y)/\partial x^2, \partial^3\varphi(x, y)/\partial x^3\}$ ,  $\{\partial^2\varphi(x, y)/\partial y^2, \partial^3\varphi(x, y)/\partial y^3\}$  –  $\{F_4, F_9; F_5, F_{10}\}$ ;

3) оставшиеся восемь фильтров – векторная комбинация (произведение) троек фильтров, упорядоченных по направлениям  $x, y$  декартовой системы координат –  $(F_1, F_4, F_9)$ ,  $(F_2, F_5, F_{10})$ .

Результатом имеем, *во-первых*, решетку  $4 \times 4$  в прямоугольной декартовой системе координат с вершинами  $F_i, i = 0, 1, \dots, 15$  (нумерация фильтров условна); *во-вторых*, матрицу  $4 \times 4$  уровня  $K_4$  и планигон (как геометрический образ матрицы). Все фильтры (исключение составляют  $F_0, F_1, F_2, F_3$ ) определены на уровне  $K_4$ , где разрешение выше в два раза по сравнению с уровнем  $K_2$ . Следовательно, композиционные ряды преобразований (2), (3) образуют две пирамиды с уровнями  $K_1, K_2, K_4$ : первая – интегральная, реализует преобразование (2); вторая – дифференциальная, реализует (3) на множестве направлений  $\{x^i\}$ . Вместе они образуют **Q-пирамиду**, а процесс их применения к объекту исследования для раскрытия его неопределенности – **Q-технология**. В силу теоремы Жордана-Гёльдера Q-пирамида и технология на ее основе универсальны и фундаментальны (теоремы оптимальности и единственности [1, 2]). Специфика такой пирамидальной технологии в том, что все преобразования (разложение, анализ, восстановление, синтез, принятие решения) реализуются «сверху-вниз», от целого к частному, отвечая тем самым главному признаку восприятия – целостности, а для зрительной системы – одномоментности восприятия. Особенность пирамидальной технологии – четность на уровнях (слоях, картах) и нечетность по числу уровней.

В приложении к системе зрительного восприятия на уровне сетчатки глаза такая пирамида есть «пирамида» трех слоев нейронов, надстроенных над «сетчаткой и растущих» против падающего луча света: внизу находятся рецепторы, вверху – ганглиозные клетки. Такое конструктивное образование – слой сетчатки как трехмерной системы. В приложении к системе зрительного восприятия на уровне зрительной коры головного мозга такая пирамида

есть пирамида нейронных клеток – простых, сложных и сверхсложных, образующих пирамидные клетки [3, 4].

### Анализ изображения

Анализ изображения предполагает два этапа информационных преобразований. Результатом первого является множество визуальных масс, упорядоченных в координатах исходного пространства  $\|m_{ij}\|$ . Результатом второго является вектор  $\mu = (\mu_0 \equiv m_0, \mu_1, \mu_2, \dots, \mu_{15})$ , где  $\mu_0 \equiv m_0$  – вершина  $Q$ -пирамиды;  $(\mu_1, \mu_2, \mu_3)$  – компоненты уровня  $K_2$ ; остальные – компоненты уровня  $K_4$ . Поскольку каждый фильтр из множества  $\{F_i\}$  «решает» уравнение в частных производных, то его решением должна быть интегральная кривая (допускающая представление касательными; теорема интерпретации [1, 2]), которая является изобразительным описанием объекта исследования (формализация проблемы остовов в компьютерных системах понимания изображений).

Все фильтры из  $\{F_i\}$ , реализуя преобразование (3), являются функциями чувствительности в ряде Тейлора и поэтому процедура их применения (но только после  $Q$ -преобразования и по всей области определения) – решение **задачи активной идентификации** (именно по этой причине рассматриваемая теория и получила свое название), выявляющей множество структурных элементов и их связей (учет, например, полутонов матрицы  $\|m_{ij}\|$  позволяет решать задачу параметрической идентификации к первоначально полученному результату структурной идентификации).

С позиций векторного анализа преобразования (3), реализуемые фильтрами ( $k$  – покрытиями), являются градиентами по своим направлениям чувствительности. В декартовой системе координат  $(x, y)$  преобразованиям фильтров  $F_1, F_2$  соответствуют компоненты градиента по направлениям  $x$  и  $y$ ; преобразованиям фильтров  $F_4, F_5$  – компоненты дивергенции, а  $F_3$  – компонента ротации в плоскости  $xOy$  с направлением в  $E^3$  [4]. Поэтому множество  $\{F_i\}$ , с позиции теоремы разложения Гельмгольца, необходимо и достаточно для анализа изображения, как векторного поля (но только после реализации  $Q$ -преобразования).

С позиций тензорного анализа и римановых пространств применение планигона, как евклидовой касательной в точке  $A_0$  плоскости к геометрически сложной поверхности, – анализ внутренней геометрии этой поверхности. Поэтому множество  $\{F_i\}$  на реализуемых преобразованиях вида  $\{\partial^k/\partial x^n \partial y^m\}$  обеспечивает анализ этой геометрии. Более того, по определению – кривая в пространстве аффинной связности называется геодезической, если всякий вектор, касательный к этой кривой в точке, остается к ней касательным при параллельном переносе вдоль этой кривой. Следовательно, множество  $\{F_i\}$  как множество на независимых преобразованиях поля Киллинга позволяет выявлять геодезические. Тем самым решена вторая проблема построения самообучающейся системы – конечное число преобразований изображения на заданном уровне разрешения, формирующих конечное число признаков на этапе их формирования.

### Алгебра описания изображения

Основные положения задачи синтеза изложены в работе [2]. Здесь выделим следующие моменты. Пусть каждому фильтру  $F_i \in \{F_i\} \equiv F$  соответствует координатно-определенный бинарный оператор  $V_i \in \{V_i\} \equiv V$ ; тогда компоненте  $\mu_i \neq 0$  вектора  $\mu$  допустимо поставить в соответствие оператор  $V_i$  либо  $\bar{V}_i$  в зависимости от знака компоненты. В результате вектору  $\mu$  ставится в соответствие подмножество операторов из  $\{V_i\}$ , имеющих аналогичную фильтрам конструкцию, но разное значение элементов матрицы ( $+1 \leftrightarrow 1$ ;  $-1 \leftrightarrow 0$ ). Задавая на множестве  $\{V_i\}$  операции умножения и сложения получаем алгебру описания изображения в двумерных булевых функциях. Такая алгебра есть алгебра синтеза изображений на планигоне, а синтезированный объект – **образ** (зрительный образ) наблюдаемого изображения.

Множество фильтров и им эквивалентных операторов, определенных на преобразованиях  $\partial/\partial x^i$ , допускает векторное представление, результатом которого являются:

1) семейство (алгебраических) полных групп  $\{P_{ni}\}$  вида  $P_{ni} = \{V_i, V_j, V_k\}$  мощности 35, где каждая группа изоморфна на сложный нейрон, а их подмножество мощности 32 – на кристаллографические группы;

2) семейство (алгебраических) замкнутых групп  $\{P_{si}\}$  вида  $P_{si} = \{V_i, V_j, V_k, V_r\}$  мощности 105, где каждая группа изоморфна на сверхсложный нейрон и образована из пары определенным образом связанных полных групп.

Семейства этих групп допускают свое использование на этапах принятия решений и понимания анализируемого изображения.

### Принятие решений в пространстве эталонов

В соответствии с изложенным пространство классов фактически есть пространство эталонных изображений, представленных в пятнадцатимерном признаковом пространстве (подробнее см. [3]). Последнее построено на базисе пятнадцати независимых градиентных преобразований  $\partial/\partial x^i$ , где  $i = 1, \dots, 15$ , образующих евклидово пространство  $E^{15}$ . В этом пространстве подмножества эталонных изображений в силу гипотезы компактности образуют классы близких эталонных изображений – точек подмножества, для которых близость необходимо должна определяться некоторым евклидовым расстоянием  $r$ . Такой подход естественен, правомерен и является стандартным подходом в теории распознавания образов. Однако он требует знания значения  $r$  и априорного знания того, что есть эталон класса. В качестве последнего должно выступать некоторое обобщенное изображение на подмножестве эталонных изображений, позволяющее успешно (желательно достоверно, что с позиций теории распознавания не обязательно – допустима некоторая доверительная вероятность положительного решения; с позиций теории активного восприятия решение проблемы узнавания должно реализоваться достоверно, либо с положительной ошибкой) решить проблему классификации – «свой-чужой». Такое обобщенное изображение называется **выработанным, обобщенным эталоном** класса эталонных изображений. Проблемы существования врожденных эталонов в теории распознавания образов даже не обсуждаются.

В рамках  $Q$ -технологии с позиций алгебры описания изображений пространством врожденных эталонов являются [3]:

1) пространство операторов  $V = \{V_i\}$  с базисом  $\mathbf{e}_i$ , где орт (репер)  $\mathbf{e}_i$  определен координатным базисом  $\{\partial/\partial x^i\}$ . Поскольку множество  $V$  имеет мощность, равную пятнадцати, то для полного представления необходимо евклидово пространство  $E^{15}$ ;

2) пространство полных групп  $P_n = \{P_{ni}\}$  со своим базисом  $\mathbf{e}_i$ . Поскольку множество  $P_n$  имеет мощность 35 (или 36, если учитывать единственную полную группу (удовлетворяющую соответствующему определению) – оператор  $V_0$ ), для полного представления необходимо евклидово пространство  $E^{35}$  (или  $E^{36}$ );

3) пространство замкнутых групп  $P_s = \{P_{si}\}$  со своим базисом  $\mathbf{e}_i$ . Поскольку множество  $P_s$  имеет мощность 105, то для полного представления необходимо евклидово пространство  $E^{105}$ .

Таким образом, имеем три типа пространств эталонов. Если каждое из них, а точнее каждое направление (в силу независимости) считать направлением поля Киллинга, то получаем три типа евклидовых пространств размерности: – пять для множества  $V$ ; – восемь для множества  $P_n$ ; – 14 для множества  $P_s$ . В этом случае:

1) каждое наблюдаемое изображение, предъявленное учителем и представленное на планигоне (как бинарной матрице) ученика, есть образ и эталонное изображение для ученика;

2) образ, сформированный на подмножестве эталонных изображений, объединенных (возможно учителем) в класс, есть выработанный, обобщенный эталон.

Три класса пространств определяют естественным образом три связанных друг с другом пространств врожденных эталонов в  $E^n$ , которые образуют «потенциальные энергетические или гравитационные ямы» и могут выступать в роли базисных на множестве  $2^{16}$ :



- 1) пространство, определенное множеством  $\{V_i\}$ ,  $-E^{15}$ ;
- 2) пространство, определенное множеством  $\{P_{ni}\}$ ,  $-E^{35}$  (без учета полной группы  $V_0$ );
- 3) пространство, определенное множеством  $\{P_{si}\}$ ,  $-E^{105}$ .

Каждое выделенное пространство имеет свой ортогональный базис:  $V = \{V_i e^i\}$ ;  $P_n = \{P_{ni} e^i\}$ ;  $P_s = \{P_{si} e^i\}$ . Поскольку каждый элемент множества по каждому пространству эталонов – тяготеющий элемент – центр класса, то любой вектор  $\mu(M_i)$ , удовлетворяющий необходимым условиям, принадлежит классу эталона. И это верно по всем трем пространствам эталонов. Поэтому классы эталонов в пространстве эталонов образуют сектора с центром в начале координат, разделенные дискриминантными функциями  $g(V)$  в пространстве  $\{V_i\}$  и  $g(P_{ni})$  в пространстве  $\{P_{ni}\}$ .

Изложенные положения оформлены в виде теории активного восприятия изображения, как объекта исследования.

**Предмет теории** – свойства и закономерности системы зрительного восприятия, представленной информационными моделями.

**Объект исследования теории** – изображение, априори неопределенное в поле зрения.

**Область применения** – анализ, синтез изображений, автоматическая классификация в условиях отсутствия (либо частичного наличия) априорной информации о наблюдаемом изображении. Полученные результаты достоверны и  $\epsilon$ -точны или «абсолютно» точны при заданной погрешности.

#### Библиографический список

1. **Утробин, В.А.** Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений: учеб. пособие / В.А. Утробин; НГТУ. – Нижний Новгород, 2001. – 234 с.
2. **Утробин, В.А.** Компьютерная обработка изображений. Анализ и синтез: учеб. пособие / В.А. Утробин; НГТУ. – Нижний Новгород, 2003. – 228 с.
3. **Утробин, В.А.** Компьютерная обработка изображений. Принятие решений в пространстве эталонов: учеб. пособие / В.А. Утробин; НГТУ. – Нижний Новгород, 2004. – 221 с.
4. **Утробин, В.А.** Физические интерпретации элементов алгебры изображения // УФН. 2004. Т. 174. № 10. С. 1089–1104.

*Дата поступления  
в редакцию 30.03.2010*

**V.A. Utrobin**

#### ELEMENTS OF THE THEORY OF ACTIVE PERCEPTION OF IMAGES

Visual perception, Substantive provisions of the theory of active perception are stated as information model of congenital mechanisms of system of visual perception of a predesign stage of images in the conditions of aprioristic uncertainty.

*Key words:* recognition of images, aprioristic uncertainty, the analysis of images, information models, stages of processing of images.