

УДК 577.3

**БИОФИЗИКА СЛОЖНЫХ СИСТЕМ.
МАТЕМАТИЧЕСКИЕ МОДЕЛИ****СИСТЕМА АНАЛИЗА ИЗОБРАЖЕНИЯ И УЗНАВАНИЕ
ОБЪЕКТОВ НА СЛОЖНОМ ФОНЕ**

Г. М. ЗЕНКИН, А. П. ПЕТРОВ

Институт проблем передачи информации АН СССР, Москва

Результаты электрофизиологических исследований зрительного анализатора свидетельствуют о том, что на начальных уровнях этой системы осуществляется разносторонний анализ изображения, проектируемого на мозаику фоторецепторов. Анализу подвергаются форма изображения, движение, цвет и т. п.

Этот анализ осуществляется элементами, имеющими размеры рецептивного поля меньше размеров всей сетчатки или всего изображения, и именно поэтому на начальных уровнях он представляет собой анализ различных «локальных» характеристик изображения. Можно сказать, что такой анализ включает в себя классификацию отдельных участков изображения, определение некоторых параметров изображения в этих участках. По терминологии, принятой в электрофизиологии, элементы, осуществляющие такой анализ, названы «детекторами».

В литературе описаны детекторы «горизонтального края» [1], детекторы «кривизны» [2], детекторы ориентированных отрезков линий [3], детекторы направления движения [4] и др. В отношении многих из них остается пока не выясненным, состоит ли их функция именно в том, что указано в их названии, как они организуются, какого рода процессы взаимодействия лежат в их основе и формируют эти свойства [5]. Тем не менее известно, что все подобные элементы, имеющие рецептивное поле относительно небольших размеров, реагируют избирательно на некоторые локальные особенности изображения.

Вся система анализа, организованная на основе таких элементов, без сомнения, призвана служить для решения определенных задач, возникающих перед человеком и животными. Однако большинство тех задач, которые приходится решать с помощью зрительного анализатора, имеет одну характерную особенность: решение зрительных задач связано с выделением, обнаружением или узнаванием целых объектов, а не отдельных их частей, целых ситуаций или обстановок, а не отдельных их элементов. Причем правильность решения этих задач, на первый взгляд, мало связана с наличием отдельных устойчивых локальных особенностей объектов, ситуаций или обстановок. Многие такие части (или даже все), вроде бы присущие этим объектам, могут сильно изменяться. Часть из них может отсутствовать. Однако в целом объект, ситуация или обстановка останутся теми же, что и раньше.

В настоящее время нет четких представлений о том, каким же образом решение задач, в которых малое значение имеет каждый отдельный элемент или часть изображения, может осуществляться системой, на первых уровнях которой стоят элементы, исследующие лишь локальные

свойства изображения. Принято считать те этапы анализа, о которых нам что-то известно, весьма и весьма предварительными и относить способность решения указанных выше задач к более высоким уровням зрительного анализатора, о механизмах работы которых известно очень мало.

Некоторые сведения о свойствах зрительного восприятия человека, относящиеся, по-видимому, к работе достаточно высоких уровней зрительного анализатора, можно почерпнуть из результатов психологических экспериментов [6—8]. Эти работы с достаточной определенностью указывают на «целостный» характер зрительного восприятия человека. Под этим прежде всего имеется в виду то, что восприятие целого объекта опережает во времени восприятие его частей. Однако конкретные механизмы, формирующие целостный характер зрительного восприятия, совершенно не ясны.

При таком положении дел может оказать большую пользу моделирование функций, выполняемых системой в обычных условиях, особенно если при этом преследуется цель воспроизведения с возможной полнотой структуры систем. В настоящей работе описывается модель системы локального анализа изображения и делается попытка решения некоторого класса задач, обладающих указанными выше особенностями, с использованием этой модели. При построении модели имелось в виду, с одной стороны, использовать элементы, похожие на те, что известны из электрофизиологии, а с другой стороны, промоделировать некоторые особенности функционирования зрительной системы. Особое внимание было обращено на умение обнаруживать и узнавать объекты на сложном фоне.

В настоящее время проведены предварительные эксперименты с моделью. Система научилась различать 25 классов изображений, включающих в себя буквы русского алфавита от А до Ф и 5 типов геометрических фигур, а также узнавать эти объекты на сложном фоне.

Описание модели

Модель, предназначенная для решения задач узнавания простых изображений букв и геометрических фигур, создана в виде программы для цифровой вычислительной машины. Основными блоками программы является блок первичного «локального» анализа изображения и блок синтеза, при помощи которого происходит обучение каждому конкретному классу изображений.

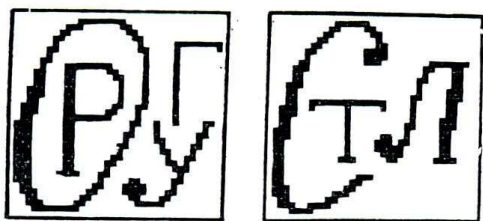


Рис. 1. Примеры изображений, подаваемых на «сетчатку»

«Сетчатка», на которую проектируется изображение, состоит из 1296 элементов (36×36), способных принимать два состояния — черное и белое. Черными являлись те элементы сетчатки, на которые падало изображение. Края сетчатки не сшивались, т. е. она являлась плоской, однако дальнейшая обработка была построена таким образом, чтобы не возникали никакие краевые эффекты. На рис. 1 приведены примеры изображений на такой сетчатке. Эти изображения и аналогичные им употреблялись в экспериментах с моделью.

Блок первичного анализа. С сетчаткой соединялось несколько типов операторов, осуществлявших локальный анализ изображения, помещенного на сетчатку. Операторы представляют собой сумматоры с порогом, имеющие один выход типа «да — нет» и 36 (6×6) входов (рецептивное поле). Друг от друга операторы одного типа не отличались, а операторы разных типов отличались рисунком связей с элементами сетчатки (рис. 2), весами входов (возбуждающий $+1$, тормозящий -1 , нейтральный

0) и порогом. Операторы каждого типа объединялись в поля размером 36×36 с поточечным соответствием поля оператора с сетчаткой. Каждый из 1296 операторов одного поля, таким образом, определял локальные свойства изображения в соответствующей точке.

При создании модели мы остановились на трех типах локальных особенностей изображения. Эти особенности можно назвать углами, концами

и узлами. Рецептивные поля

были получены в соответствующих типов операторов, а также порог их срабатывания в процессе отбора. Выбор операторов локальных особенностей производился только среди таких, рецептивное поле которых имело четыре оси симметрии. Такие поля генерировала отдельная подпрограмма. Отбор осуществляла другая подпрограмма, которая оценивала работу различных операторов на наборе контрольных картинок. Эта оценка производилась с помощью трех типов масок, на каждой из которых были указаны места расположения соответствующих локальных особенностей изображения (концов, углов и узлов) на контрольных картинках. Требовалось, чтобы оператор соответствующего типа возбуждался только в местах, указанных на маске. Сравнением поля оператора с полем маски и определялось качество оператора. Примеры рецептивных полей отобранных операторов приведены на рис. 2, а, б и в.

По некоторым причинам, в частности из-за сравнительно простого устройства рецептивного поля и неполной симметрии его, один оператор не мог выделять все локальные особенности своего типа. Однако несколько операторов, отобранных для выделения локальных свойств одного типа, справлялись с этой задачей вполне удовлетворительно. Было исследовано несколько способов объединения операторов одного типа, и самым простым и надежным оказалось логическое сложение их полей.

Такое простое устройство рецептивных полей операторов локальных особенностей потребовало обязательного предварительного выделения контуров изображения. Поэтому между слоем рецепторов (сетчаткой) и операторами локального анализа в модели помещены операторы выделения контура изображения. Именно с полем контура имели связи операторы локального анализа.

В модели выделялись два типа контуров — внутренний и внешний. Смысл этих операций виден на рис. 3, Б и В. Выделение внешнего контура производилось оператором, имеющим рецептивное поле размером 3×3 элемента с центральным возбуждающим входом, имеющим вес $+8$, и периферическими тормозящими входами с весом -1 . Порог оператора равен 1.

Выделение внутреннего контура изображения производится последовательным применением нескольких операторов, построенных по типу оператора внешнего контура. Обе операции превращают изображение в контурный рисунок, линии которого имеют толщину в один элемент. Дальнейшая обработка полученных контуров происходит до некоторого этапа одинаково и независимо.

Кроме операторов концов, углов и узлов для локального анализа изображения в модели использовались операторы отрезков линий четы-

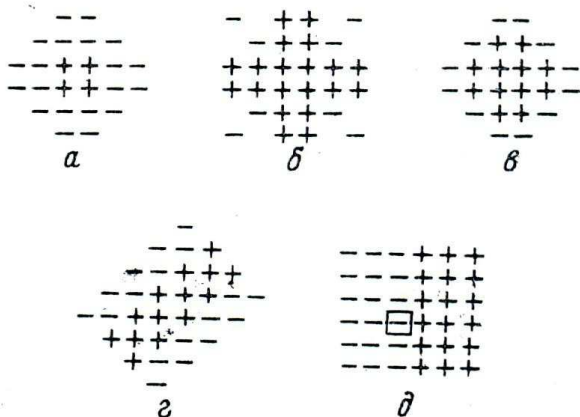


Рис. 2. Рецептивные поля операторов анализа

а — оператор концов; б — оператор узлов; в — оператор углов; г — оператор наклонно ориентированных отрезков; д — оператор ориентированных локальных особенностей

рех ориентаций. Операторы ориентированных линий устроены так же, как и операторы локальных особенностей. Рецептивное поле одного из них показано на рис. 2, г. Порог срабатывания этих операторов был подобран в эксперименте.

Следующие этапы локального анализа изображения проводят классификацию концов, углов и узлов в зависимости от их ориентации на

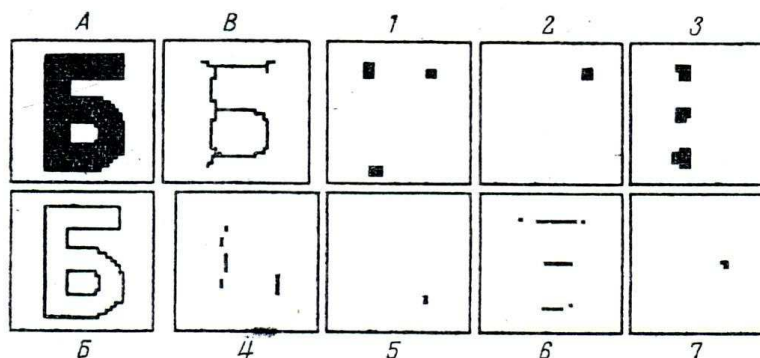


Рис. 3. Преобразование изображения в блоке анализа
 А — изображение на сетчатке; Б — внешний контур этого изображения; В — внутренний контур; 1—7 — примеры полей анализа изображения В: 1 — поле концов; 2 — углов; 3 — узлов; 4—7 — поля ориентированных линий

сетчатке. При этом также различаются четыре ориентации. Таким образом, получается по четыре поля ориентированных локальных особенностей каждого из трех типов. В модели эта классификация производится операторами с рецептивными полями (рис. 2, д). Все входы этого оператора соединены с полем контура, но срабатывает оператор только при попадании на соответствующую особенность изображения, т. е. можно считать, что оператор ориентированных концов (углов, узлов) имеет еще один вход, который соединен с полем концов (углов, узлов).

После простой операции удаления из контура изображения всех выделенных ранее частей (концов, углов, узлов и ориентированных линий) остается поле, которое названо полем кривых линий. С помощью операторов ориентированных линий с пониженным порогом производился анализ этого поля и получалось еще четыре поля ориентированных кривых линий. На этом локальный анализ изображения оканчивался. Всего, таким образом, получается 24 поля анализа, не считая самого изображения и его контура*. Пример проведенного в эксперименте анализа одной из букв дан на рис. 3, 1—7.

Можно видеть, что локальный анализ представляет собой классификацию всех точек изображения, учитывающую их взаимное расположение в пределах рецептивного поля оператора. Именно поэтому он является локальным. Для решения интересующего нас класса задач мы перешли к формированию нелокальных операторов, анализирующих взаимное расположение различных частей изображения. Эксперименты показали, что для решения тех задач, которые были предложены системе, достаточно было сформировать лишь один нелокальный оператор на каждый класс объектов. Эти операторы были названы гештальт-операторами (ГШ). Таким образом, обучение системы заключалось в синтезе ГШ-оператора, представлявшего собой память (эталон) одного класса.

Описание ГШ-оператора. ГШ-оператор можно представить в виде дизъюнкции трех операторов, входы которых соединены со всеми 24 полями анализа. На рис. 4 схематически показано устройство одного из трех операторов, входящих в ГШ-оператор. На каждом поле анализа он

* Опыты показали, что достаточно использовать поле одного, например, внутреннего контура.

имеет по сто входов, расположенных в виде квадрата 10×10 элементов. Центры всех таких квадратов на разных полях анализа соответствуют одной и той же точке сетчатки.

Из рис. 4 видно, что оператор состоит из сумматора Σ с порогом, каждый вход которого соединен с полем анализа через промежуточный сумматор (ПС, рис. 4). Всего сумматор Σ имеет 600 входов (по 25 на каждом поле анализа), веса которых формируются в процессе обучения. Промежуточные сумматоры имеют по 4 входа (2×2) и возбуждаются, если хоть один из входов возбужден. Рецептивные поля промежуточных сумматоров не перекрываются.

Два других оператора, входящих в ГШ-оператор, имеют точно такой же сумматор Σ . Отличаются они друг от друга и от описанного выше лишь числом входов промежуточных сумматоров, имеющих по 16 (4×4) и 64 (8×8) входов соответственно. С помощью слоев промежуточных сумматоров производится подобное преобразование полей анализа. Этим осуществляется перебор по различным размерам изображения и достигается инвариантность ответа ГШ-оператора по отношению к преобразованию подобия. ГШ-операторы каждого класса, так же как операторы анализа, объединяются в поле. Этим достигалась инвариантность ответа системы относительно преобразования сдвига изображения по сетчатке.

Синтез гештальт-оператора. Для синтеза ГШ-оператора каждого класса достаточно, как следует из его схемы, сформировать веса входов одного из трех его сумматоров Σ . Вес каждого из 600 входов сумматора Σ мог принимать значение от -15 до $+15$. На этапе синтеза системе предъявляли изображения без шума. Алгоритм формирования сумматоров Σ таков.

Системе предъявили последовательно по одному представителю каждого класса с указанием того, к какому классу он относится. После обработки изображения и получения 24 полей анализа все поля уменьшались в 2, 4 или 8 раз так, чтобы все элементы изображения попали в квадрат размером 5×5 элементов, т. е. уместились в поле входов сумматора Σ на каждом поле анализа. Затем из всех полей анализа вырезались такие квадраты с центром в точке, соответствующей заранее определенному «центру» изображения на сетчатке. После этого каждое изображение, превращенное в результате преобразования в 600 (24 поля размером 5×5) элементов, принимающих значение 0 или 1, поступало на входы сумматоров Σ , которые соединялись с полями анализа непосредственно без промежуточных сумматоров. Суммируя веса возбужденных входов сумматора Σ , получали его ответ в виде числа. Программа выводила на печать ответы сумматоров Σ всех классов.

В зависимости от этих ответов веса тех входов (соответствующего сумматора Σ), которые были возбуждены, увеличивались на единицу, если система делала ошибку, или оставались неизменными, если система правильно узнавала объект, т. е. ответ соответствующего сумматора Σ был больше остальных на некоторую заранее установленную величину. Обычно в начале обучения веса не уменьшались — происходила только суммация.

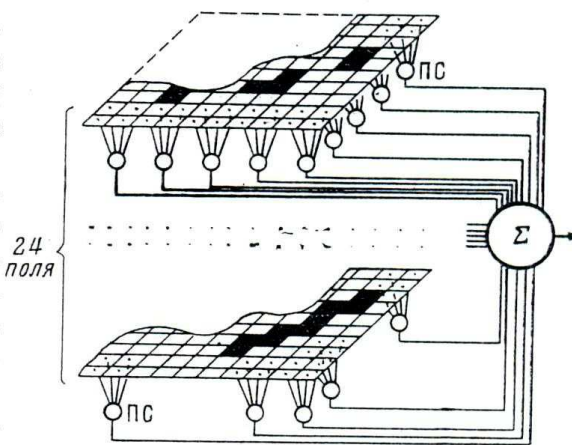


Рис. 4. Схема к описанию устройства гештальт-оператора

ПС — промежуточные сумматоры

После достаточного обучения, когда в сумматоры Σ было записано по 5—6 представителей каждого класса, ошибки системы состояли в том, что она путала лишь похожие изображения, например некоторые Б и В, С и Е, Г и Т и т. д. В таком случае в процессе обучения осуществлялась так называемая «дифференцировка», которая заключалась в том, что одновременно с увеличением весов своего сумматора Σ происходило уменьшение весов сумматора Σ того класса, к которому данное изображение было ошибочно отнесено.

Процесс обучения оканчивался, когда система не делала ни одной ошибки на материале обучения. Устройство ГШ-операторов и алгоритм обучения позволяли последовательно доучивать систему узнаванию новых классов изображения.

Результаты экспериментов

Основной целью описанных здесь предварительных экспериментов было выяснение возможности программы узнавать объекты на сложном фоне. Проведенные с моделью эксперименты разделяются на две основ-

ные группы. Первую из них составляют опыты, в которых проверялась способность системы к обобщению и различению при наличии большого числа классов объектов. Для этого было проведено обучение системы (формирование ГШ-операторов) различению 20 классов букв русского алфавита от А до Ф и 5 классов геометрических фигур, предъявляемых без шума. Как видно на рис. 5, изображения букв, предъявлявшихся системе, были очень разнообразны. В одном из таких опытов было отобрано по 15 наиболее разнообразных изображений букв, которые затем предъявлялись системе для обучения. После формирования ГШ-операторов для 20 классов был проведен опыт по узнаванию 700 (35×20) букв. На этом материале система ошибочно классифицировала восемь изображений букв.

Ошибки, сделанные системой, имели указанный выше характер. Наблюдалось ошибочное отнесение букв А к клас-



Рис. 5. Изображения букв 20 классов, предлагавшихся системе при обучении

сам Н, Л, Р; Б и классу В; Л к П и др., т. е. система путала между собой изображения похожих для человека букв (рис. 5).

Во второй группе экспериментов исследовалась способность системы, обученной 20 классам букв, узнавать объекты этих классов на сложном фоне. В одной из серий этих опытов фон в виде отдельных точек генерировался случайным образом. Рисунок 6, I показывает примеры изображений букв, на которых система делала около 20% ошибок. Ошибки имели тот же характер.

В другой серии опытов модели предъявлялись изображения, примеры которых приведены на рис. 6, II (и на рис. 1). В случаях (всего около 30), когда изображение одной буквы было перечеркнуто линией, система правильно узнала все изображения. Большой интерес представляла работа программы при предъявлении ей сразу нескольких букв на сетчатке, иногда частично перекрывавшихся. Анализ полей ответов ГШ-операторов, которые программа выпечатывала, показал, что почти во всех случаях система правильно указала буквы и места их расположения на поле. Однако в ряде случаев система указывала на наличие букв там, где их не было. Эти ошибки связаны в основном с тем, что сочетание нескольких линий, принадлежавших разным буквам, походило на изображение одной из букв 20 классов.

Обсуждение

Укажем характерные черты, которыми, по нашему мнению, обладают задачи, решаемые зрительным анализатором человека и животных.

Прежде всего, рассматриваемые задачи обладают своеобразной структурой, выделяющей их из задач распознавания образов, которые порой ставятся в слишком общих терминах. Признание этой структуры приводит к важным ограничениям на возможную модель зрительной системы.

Если придерживаться принятой терминологии [9], можно сказать, что зрительный анализатор решает задачу распознавания образов, в которой число классов настолько велико, что бессмысленно заниматься описанием границы между ними. Кроме того, это бесполезно и потому, что существует еще один класс, который содержит все состояния сетчатки, отвечающие произвольным не реализующимся в обычных условиях распределениям освещенности. В многообразии этих распределений изображения реальных объектов не занимают практически никакого объема и представляют собой многообразия значительно меньших размерностей. Указание на это можно найти в работе Файна [10].

Кроме того, какие бы мы ни выбрали достаточно разумные параметры изображений, обычно приходим к выводу, что объединение объектов одного класса является серьезной задачей, так как изменчивость объектов очень велика. Это приводит, вследствие сильной взаимной зависимости между параметрами (у реальных объектов), к очень сложной характеристической функции каждого класса. А именно эти функции нам и следует построить для решений задачи.

В модели эти характеристические функции классов реализованы в виде гештальт-операторов. Одной из важных черт системы ГШ-операторов является инвариантность ответа по отношению к преобразованиям сдвига, поворота** и подобия.

Необходимость такой инвариантности следует из того, что для правильного поведения во внешней среде положение и размеры изображения относительно сетчатки не имеют значения и объективного смысла. Имеют смысл лишь пространственные отношения объектов между собой.

** В проведенных экспериментах механизмы, приводящие к полной инвариантности по отношению к повороту, не использовались и по этой причине здесь не описаны.



Рис. 6

Говоря о связи модели с физиологией и психологией, можно утверждать, что этапы первичного анализа и устройство операторов анализа, без сомнения, опираются на физиологические данные. Этого нельзя сказать о блоке синтеза. Элементы, подобные ГШ-операторам, в физиологии не описаны. Однако результаты экспериментов с программой позволяют говорить, что на уровне синтеза ГШ-операторов она моделирует некоторые важные функции зрительной системы. Это особенно ясно выступает при узнавании фигур на сложном фоне. Такая способность несомненно достигнута благодаря блоку синтеза.

Функционирование модели зрительного анализатора, имеющей первыми этажами описанную в работе систему, будет обладать интересной особенностью, которую можно описать как целостность восприятия объектов. Под этим мы имеем в виду, что узнавание объекта как целого происходит прежде, чем производится анализ частей, из которых объект состоит. Это парадоксальное, на первый взгляд, высказывание разъясняется следующим образом.

Когда речь идет о частях целого объекта, то это есть его функциональные части, а не формальные части изображения. Именно в этом смысле нижнюю часть букв А можно назвать концами, хотя в геометрическом смысле это может быть и не так. В модели синтез ГШ-операторов опирается на анализ формальных частей изображения, а не функциональных частей объекта. Первичный анализ изображения разбивает его на части, которые находятся, однако, в сложном отношении с функциональными частями этого же объекта. Затем система синтезирует из этих частей образ объекта до того, как она осуществляет выделение его функциональных частей. Именно поэтому система может отвечать на вопрос, какая перед ней буква, «не разглядев», какого типа концы у этой буквы и т. п. Аналогично можно узнать кошку, не заметив, два у нее уха или одно, как расположен хвост и т. п.

После узнавания объекта как целого возможен анализ его функциональных частей с целью решения других более сложных задач. В описанной здесь модели анализ частей изображения, как частей целого объекта, не воспроизводится.

Итак, мы представляем себе зрительную систему как систему, имеющую три основных этапа работы при восприятии: 1) первичный анализ изображения, или рецепторный анализ; 2) этап синтеза, т. е. формирование гештальт-ответа; 3) анализ функциональных частей объекта, т. е. рецепторный анализ.

Между этими этапами и структурами, их выполняющими, существуют очень сложные связи и взаимоотношения (например, возврат к первичному анализу после рецепторного и т. п.), что в значительной мере может затруднить их выделение в экспериментах с животными и людьми.

Способность обнаруживать объекты на сложном фоне получена после обучения системы на рисунках без фона. Это обстоятельство не является принципиальным отличием от обучения животных в реальной обстановке. Легко видеть, что движение объекта относительно объектов фона делает возможным осуществление обучения на «чистых» изображениях. Для этого, по-видимому, необходимо наличие детекторов движения и механизма слежения.

В заключение отметим, что описанная модель очень хорошо приспособлена к выполнению ее в виде аналогового устройства. При этом она имела бы всего 4—5 последовательных слоев пороговых элементов, что, по-видимому, мало отличается от числа этапов зрительной системы мозга. Вынужденная же последовательность операций при реализации ее в виде программы для ЦВМ требует довольно большого количества машинного времени.

Выводы

1. Предлагается система, предназначенная для узнавания зрительных образов, состоящая из блока первичного анализа изображения и блока синтеза нелокальных операторов, названных гештальт-операторами.

2. Построенная система в виде программы для ЦВМ в настоящее время обучена различению 25 классов изображений, включающих изображения 20 букв русского алфавита и 5 классов геометрических фигур.

3. Система успешно справляется с задачей узнавания объектов этих классов, предъявляемых на сложном фоне.

4. Рассматривается место предложенной системы в более полной модели зрительного анализатора.

* * *

Авторы выражают глубокую признательность А. С. Кронроду и В. Л. Арлазарову за предоставленную возможность проводить эксперименты на ЭВМ.

ЛИТЕРАТУРА

1. Maturana H. R. a. Frenk S., Science, **142**, 977, 1963.
2. Maturana H. R., Lettvin J. Y., McCulloch W. S. a. Pitts W. H., J. Gen. Physiol., **43**, 129, 1960.
3. Hubel D. H. a. Wiesel T. N., J. Physiol., **160**, 106, 1962.
4. Barlow H. B., Hill R. M. a. Levick W. R., J. Physiol., **173**, 377, 1964.
5. Barlow H. B., Levick W. R., J. Physiol., **178**, 477, 1965.
6. Вудвортс Р. Экспериментальная психология, ИЛ, М., 1950.
7. Дембовский Я., Психология животных, ИЛ, М., 1959.
8. Глезер В. Д., Механизмы опознания зрительных образов, Изд-во «Наука», М., 1966.
9. Сочивко В. П., Распознавание образов при помощи вычислительных машин. Итоги науки: теория вероятности и математическая статистика, 1964, М., 1966.
10. Файн В. С., Изв. АН СССР, Техн. киб., № 2, 1966.

Поступила в редакцию
24.XI.1966