На правах рукописи

АКСЁНОВ Сергей Владимирович

МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ РАСПОЗНАВАНИЯ МНОГОЭЛЕМЕНТНЫХ ЗРИТЕЛЬНЫХ СЦЕН С ИСПОЛЬЗОВАНИЕМ ИЕРАРХИЧЕСКИХ НЕЙРОННЫХ СЕТЕЙ

Специальность: 05.13.11 – *«Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»*

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата технических наук

Томск – 2008

Работа выполнена в государственном образовательном учреждении высшего профессионального образования Томский политехнический университет

Научный руководитель: ***НОВОСЕЛЬЦЕВ Виталий Борисович***, доктор физико-математических наук, доцент

Официальные оппоненты: ***МАТРОСОВА Анжела Юрьевна***, доктор технических наук, профессор
КОРИКОВ Анатолий Михайлович, доктор технических наук, профессор

Ведущая организация: ***Алтайский государственный университет, г.Барнаул***

Защита состоится «29» декабря 2008 г. в 14:30 на заседании совета по защите докторских и кандидатских диссертаций **Д212.269.06** при Томском политехническом университете по адресу: 634034, Томск, ул.Советская, 84, Институт «Кибернетический центр» ТПУ.

С диссертацией можно ознакомиться в научной библиотеке Томского политехнического университета по адресу: 634050, г.Томск, ул.Белинского, 55.

Автореферат разослан «28» ноября 2008г.

Ученый секретарь
совета по защите докторских
и кандидатских диссертаций
к.т.н., доцент ***М.А.Сонькин***

1.ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИИ

Актуальность темы

Существующие в настоящее время системы распознавания на базе искусственных нейронных сетей (ИНС) обладают значительным количеством недостатков. При разработке комплексов, решающих проблему интерпретации изображений, особый упор делается на узнавание определенных групп образов. Таких примеров, достаточно много. Однако, поиск в доступных источниках информации о системах, способных к саморазвитию (к «расширению кругозора»), приводит к крайне скудным результатам. Это объясняется высокой сложностью проектирования и реализации проектов такого класса. Ведь на самом деле реальный процесс распознавания, протекающий в самой совершенной системе анализа – человеческом мозге, не состоит только в том, чтобы проверить сходство анализируемого объекта с запомненным эталоном. Это достаточно сложное взаимодействие между различными подсистемами мозга. Ошибочно также считать, что сигналы от сетчатки до терминальных корковых центров, принимающих решение, распространяются прямолинейно. Распознавание – это итеративный процесс, в котором до принятия окончательного решения не только входной образ сравниваются с эталоном, но и генерируются гипотезы по классификации объекта. Столкновение соответствующих потоков нейронных импульсов приводит к дальнейшему уточнению характеристик образа, до тех пор, пока не произойдет согласования между двумя потоками.

Достаточно часто в процессе классификации случается так, что наш мозг не содержит точного представления эталона, однако это не мешает ему правильно найти результат. Другим немаловажным аспектом является выделение существенных фрагментов сцены для их интерпретации, т.е. должна производиться селекция образов, расположенных на сцене. В противном случае, анализируется вся сцена, со всеми вытекающими последствиями полного перебора пространства решений. Существующие реализации таких комплексов в качестве входных данных требуют практически полностью очищенные от шумов сцены, а при даже незначительных зашумлениях дают неверные результаты.

Таким образом, актуальность исследований обосновывается необходимостью разработки математического аппарата новых программных архитектур, алгоритмов функционирования и обучения ИНС для обеспечения более надежных результатов распознавания сложных объектов, находящихся на произвольном фоне. При этом основные усилия направлены на разработку интеллектуальной составляющей иерархической ИНС, позволяющей программному комплексу анализировать только существенные компоненты входного вектора.

Теоретико-методологической основой диссертационной работы явились исследования в области компьютерного зрения, нашедшие отражение в трудах Хьюбела Д., Визела Т., Фукушимы К., Гроссберга С., ЛеКуна Й., Ризенхубера М., Поггио Т., Карпентер Г.А., Ли Ж., Цоцос Д.К., Итти Л., Бейнке С.

Целью является разработка математической модели иерархической ИНС и программной реализации для решения задач распознавания образов, сгруппированных в произвольную сцену.

В связи с этим в диссертационной работе решаются следующие **задачи**:

1. Разработка математического описания архитектуры иерархической ИНС, позволяющей представить образ объекта из сцены на разных уровнях обобщенности.
2. Проектирование алгоритма настройки подсистемы распознавания иерархической ИНС.
3. Разработка механизма совместного функционирования подсистем внимания и распознавания.
4. Программная реализация разработанных алгоритмов настройки и функционирования иерархической ИНС.

Методы исследования

В работе использован инструментарий теории алгоритмов, методы объектно-ориентированного проектирования и программирования, теории нелинейной оптимизации, теории распознавания образов, теории функционирования биологических нейронных систем.

Научная новизна

- Разработана математическая модель системы распознавания связанных фрагментов образов на основе взаимодействия подсистем внимания и распознавания.
- Предложен оригинальный алгоритм настройки иерархической ИНС с использованием механизмов обучения с учителем и без учителя.
- Разработан алгоритм функционирования иерархической ИНС, фокусирующийся на значимых характеристиках образа.
- Предложен алгоритм выделения контрастных связанных фрагментов сцен на базе процессов сегментации и восходящего внимания.
- Реализован программный комплекс, выполняющий поиск существенных фрагментов сцены с их последующим анализом.

Практическая ценность работы заключается в следующем:

- Предложенные математические модели и алгоритмы обработки визуальной информации для анализа сцен позволяют эффективно решать задачи распознавания зашумленных образов, расположенных на сложном фоне. Предложен более эффективный по сравнению с существующими моделями иерархических нейронных сетей способ определения класса объекта, расположенного на сцене. Разработанные алгоритмы и их программная реализация внедрены в Научно-исследовательском институте оптики и атмосферы СО РАН.
- Разработанная автором подсистема внимания позволяет оптимизировать вычислительные ресурсы, направив их исключительно на анализ существенных фрагментов сцены.
- Предложенный в работе механизм взаимодействия подсистем внимания и распознавания может использоваться отдельно от всего комплекса сторонними

разработчиками в собственных исследованиях обработки визуализированных сцен.

- Отдельные модули разработанного программного комплекса имеют самостоятельное значение и могут использоваться в автономном режиме, что делает данный комплекс (подход) универсальным инструментом построения архитектур и обучения нейронных сетей для решения различных прикладных задач нейрокомпьютинга, таких как классификация, кластеризация, аппроксимация и управление.

Апробация работ

Результаты работы докладывались на: III Всероссийской научно-практической конференции «Научная сессия ТУСУР–2004». – Томск, 2004; VI Всероссийской конференции «Молодежь и современные информационные технологии». – Томск, 2006; V Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых. «Молодежь и современные информационные технологии». – Томск, 2007; XLV Международной научной студенческой конференции "Студент и научно-технический прогресс". – Новосибирск, 2007; XIII Международной научно-практической конференции студентов и молодых ученых «Современные техника и технологии». – Томск, 2007; VI Всероссийской научно-практической конференции «Инновационные недра Кузбасса. IT-технологии». – Кемерово, 2007; VI Всероссийской конференции «Молодежь и современные информационные технологии». – Томск, 2008; Всероссийской конференции «Информационные технологии в авиационной и космической технике – 2008». – Москва, 2008; XLVI Международной научной студенческой конференции «Студент и научно-технический прогресс». – Новосибирск, 2008; XIII Всероссийской научно-технической конференции «Новые информационные технологии в научных исследованиях и образовании». – Рязань, 2008; III Международной научно-технической конференции «Инфокоммуникационные технологии в науке, производстве и образовании». – Кисловодск, 2008; XV Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов», – Москва, 2008; Всероссийская научно-практическая конференция Информационные технологии в профессиональной деятельности и научной работе – Йошкар-Ола, 2008; XIV Международной научно-практической конференции студентов и молодых ученых «Современные техника и технологии». – Томск, 2008.

Основные положения, выносимые на защиту

1. Математическая модель иерархической ИНС для анализа сложных изображений, учитывающая взаимодействие процессов фиксации внимания и собственно распознавания.
2. Алгоритмы обучения и функционирования иерархической ИНС предложенной структуры.
3. Алгоритм выделения значимых фрагментов сцены подсистемой внимания.
4. Программная реализация интерпретации сцен на базе взаимодействия подсистем внимания и распознавания.

Публикации

Основные положения диссертационной работы изложены в 14 работах, в том числе 1 монография. Монография награждена дипломом Всероссийского конкурса Фонда развития отечественного образования «За лучшую научную книгу 2006 года». Личный вклад автора в каждой работе составляет 50%–100%.

Личный вклад автора

Основные результаты диссертационной работы получены автором лично. Программный комплекс «HNNMeister 1.0» для проектирования архитектур, настройки и использования иерархических нейронных сетей разработан автором лично.

Внедрение результатов

Результаты работы используются в учебном процессе на кафедре Оптимизации систем управления ТПУ, на кафедре Программирования ТГУ, внедрены в Институте оптики атмосферы СО РАН (г.Томск).

Структура и объем диссертации

Диссертация состоит из введения, четырех глав, заключения, списка литературы (136 источников). Материал изложен на 154 страницах, содержит 10 таблиц, 57 рисунков.

2. СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обосновывается актуальность темы диссертационной работы, определяются предмет и цели исследования, формулируются задачи исследования, а также указываются возможные варианты использования полученных результатов.

В первой главе систематизированы распространенные подходы и методы нейрокомпьютерных вычислений, применяемые для обработки изображений. Представлен достаточно полный, на наш взгляд, аналитический обзор современных направлений исследований в области нейросетевого зрения.

Современная теория нейросетевых вычислений предлагает большое число моделей и алгоритмов НС, в той или иной степени позволяющих распознать образ на сцене. Однако большинство моделей имели значительные ограничения к представлению образа. Другими словами, такие системы могли выполнить распознавание только при условии минимальных шумов, отсутствии трансформации анализируемого объекта, расположенного на «белой» сцене. Кроме этого, образ требовал определенного масштабирования. В конечном счете, подобные ограничения привели к сужению области распространения таких систем. Как показано в материалах главы, для создания систем, обладающих большими возможностями, потребовалось ввести определенную специализацию частей ИНС, задать соответствующую архитектуру, увеличить количество типов нейронов и синаптических связей, используемых внутри модели. Ввиду такого усложнения большинство исследователей сосредоточили свои усилия в относительно узких областях теории распознавания, таких как анализ рецептивных полей или конкурентное взаимодействие нейронов. Однако исследования, ставившие своей целью конструирование универсальной

системы распознавания объектов, в результате создали огромные и в значительной степени избыточные архитектуры ИНС. Процесс анализа образа происходил большей частью перебором множества комбинаций соседних нейронов. Таким образом, к главным недостаткам существующих подходов можно отнести отсутствие интеллектуализации (направленного и осмысленного поиска) моделей. Другой недостаток – неассоциативный характер хранения образов, что самым негативным образом сказывается на распознавании.

Исходя из анализа доступных источников, сделано заключение, что предпринимаемые попытки осознать процесс распознавания являются скорее количественным, чем качественным путем решения.

Во второй главе рассматривается предлагаемая математическая модель подсистемы распознавания для выделения существенных фрагментов сцен. Составляющая ядро модели иерархическая ИНС, уровни которой представлены на рис.1, состоит из нескольких слоев, каждый из которых включает группу двумерных плоскостей нейронов. Нейроны, образующие плоскости в слоях, будем называть клетками характеристик (или свойств).

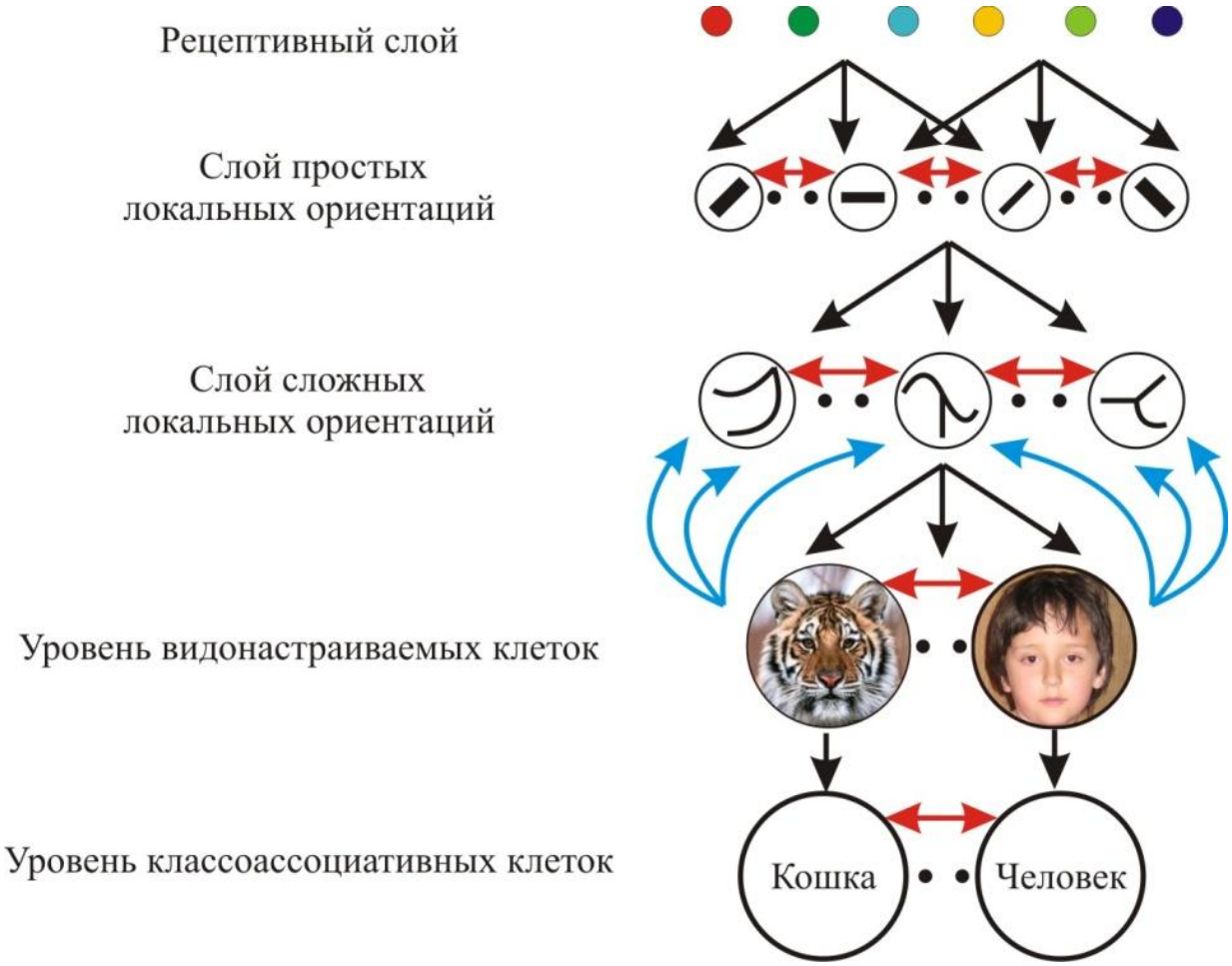


Рис. 1. Уровни иерархической ИНС

Предлагаемое представление дает возможность анализа фрагмента на нескольких уровнях абстракции. С повышением уровня происходит увеличение числа плоскостей с одновременным уменьшением пространственного

разрешения. Предложенная организация нейронов в двумерные плоскости в первых трех уровнях, изображенных на рис. 1., и MAX-слоях (дополнительный подуровень в слоях простых и сложных локальных ориентаций), мотивируется тем, что близкорасположенные фрагменты коррелируют в большей степени, нежели более удаленные друг от друга области. Информация о базовых функциях, реализуемых на каждом уровне, приведена в табл. 1.

Клетки простых локальных ориентаций, или S-клетки, (фрагментов прямых линий) обладают рецептивным полем переменного размера, что позволяет выполнить нормирование образа для последующих уровней ИНС и обеспечить в определенной степени инвариантность к масштабированию образов. Расположение нейронов с рецептивным полем переменного размера внутри слоёв простых и сложных характеристик позволяет решить две задачи: во-первых, производится оценка входного вектора нейронами, обладающими различными рецептивными полями, и, во-вторых, плоскости одного слоя включаются в латеральное соревнование. Последнее позволяет более точно определить клетки ИНС с синаптическими соединениями, настроенными на соответствующие компоненты, попавшие в рецептивное поле.

Табл.1. Функции уровней иерархической ИНС

Наименование уровня	Функции
Рецептивный уровень	Получение фрагмента сцены и хранение его полутонового представления
Слой простых локальных ориентаций	Выделение простых локальных ориентаций посредством рецептивных полей на базе дифференциального фильтра Гаусса или фильтра Габора
MAX-слой	Нахождение максимального выхода нейронов, расположенных на одной позиции в разных плоскостях, с одинаковой ориентированностью, но обладающих разной толщиной
Слой сложных характеристик	Выделение сложных характеристик образа с помощью комбинации простых характеристик предыдущего этапа
Ассоциативная среда	Управление генерацией гипотез о природе образа
Уровень видонастра-иваемых клеток	Представление образов, основанное на их визуальных характеристиках
Уровень восприятия классов	Представление классов образов

В S-клетках с переменным рецептивным полем ИНС, производится выделение значимых компонент для полутонового представления входного изображения, расположенного в рецептивном слое. К значимым компонентам образа, обрабатываемым клетками, относятся части линий, определенных ориентаций: $0, \pi/4, \pi/2, 3\pi/4$. Каждая ориентированная значимая компонента представлена

в нескольких вариантах. Другими словами, клетки анализируют не только ориентацию, но и толщину линии полутонового представления в рецептивном слое. Это делается для уменьшения зависимости качества распознавания от колебаний границ между фоном и анализируемым образом (границ между участками образа). Извлечение локальных характеристик этими плоскостями осуществляется посредством одного из двух фильтров: дифференциального фильтра Гаусса или фильтра Габора.

Размеры рецептивного поля для S-клеток рассчитываются согласно выражению:

$$h = w = max(H_2\ div\ H_1,\ W_2\ div\ W_1),$$

где h, w – соответственно ширина и длина рецептивного поля, H_1, W_1 размеры плоскости, содержащей S-клетку, а H_2, W_2 – размеры обрабатываемого фрагмента на рецептивном уровне. Значения синаптических весов устанавливаются после получения параметров h и w.

Для уменьшения избыточности данных, являющейся следствием использования групп плоскостей с равным параметром ориентации, но с отличающимся значением ширины локальной характеристики используется подход связывания выходных сигналов такой группы. Такое группирование производится в MAX-плоскостях, совпадающих по размерам с ассоциированными S-плоскостями. Каждый нейрон MAX-плоскости, расположенный на позиции ν, получает сигналы состояния всех нейронов, входящих в одну группу и расположенных на этой же позиции ν, и находит среди них максимальный отклик. Значение максимального выхода афферента и становится выходом MAX-нейрона ν:

$$U(MAX, \nu, \zeta) = \max_{k \in A_c}(U(S, \nu, k)),$$

где ζ – номер плоскости в MAX-слое, $U(S, \nu, k)$ – выход S-клетки, попавшей в область связи нейрона ν, A_c – S-плоскости, расположенные на одном уровне обладающие клетками с равным значением ориентации.

Уровень сложных характеристик, или C-слой, обладает подуровнем сложных свойств, являющихся комбинацией простых характеристик, выделенных на раннем этапе и сгруппированных в MAX-плоскостях. Для свертки простых характеристик применяется метод Фукушимы, основанный на совместном использовании возбуждающих и тормозящих сигналов. Выход клетки сложных характеристик:

$$U(C, \nu, \zeta) = r_\nu \times \varphi\left[\frac{1 + \sum_{n=1}^{K_{MAX}} \sum_{\mu \in A_\nu} a(C, \xi, \mu, \zeta) \cdot U(MAX, \nu + \mu, n)}{1 + \frac{r_\nu}{r_\nu + 1} \cdot b(C, \zeta) \cdot U(V, \nu)} - 1\right]$$

где A_ν – область связи соответствующей C-клетки, r_ν – параметр селективности (степень близости входного сигнала и сигнала, которому были научены синапсы клетки), $\varphi(\ldots)$ –порогово-линейная функция,

$$U(V, \xi) = \sqrt{\sum_{n=1}^{K_{MAX}} \sum_{\mu \in A_\nu} d(V, \xi) \cdot U^2(MAX, \nu + \mu, n)}\ ,$$

здесь K_{MAX} – количество плоскостей в MAX слое.

После определения нейрона-победителя в каждой области гиперколонарного торможения производится настройка его весов:

$$\Delta a(C,\xi,\mu,\zeta) = q \cdot d(V,\mu) \cdot U(MAX, v+\mu, n),$$
$$\Delta b(C,\zeta) = q \cdot U(V,v),$$

где q – коэффициент скорости обучения.

Предпоследний уровень модели – уровень *видонастраиваемых нейронов* (**ВНН**) служит для запоминания визуального представления целых объектов. Компоненты вектора весов каждой видонастраиваемой клетки служат для указания вклада сложных характеристик для каждого запомненного класса. Решение задачи поиска визуального представления сводится к нахождению таких ВНН, для которых расстояние компонент весового вектора до выходов С-слоя минимально.

Активность фрагмента С-слоя позволяет предположить, что существует вероятность принадлежности входного фрагмента к одному из запомненных образов. Существуют три стратегии поиска образа: восходящая, в которой процессом управляют выходы С-слоя, нисходящая, контроль осуществляется ВНН, и комбинированная, сочетающая предыдущие. Перед запуском процедуры поиска фиксируется значения D_{max} – минимально возможного отклонения сегмента от представления образа, при котором отвергается гипотеза о принадлежности сегмента к указанному виду. R – первоначальные окрестности С-клеток, инициирующих поиск, r – шаг увеличения окрестности, $U_{min}(C)$ – минимальная активность сложной клетки, при которой последняя запускает поиск, w_{min} – минимальное значение веса синапса ВНН, инициирующего процедуру восходящего поиска. Необходимо отметить, что начальные шаги всех стратегий поиска совпадают и выглядят следующим образом:

Шаг.1 Активация сети (здесь предполагается, что настройка сложных клеток уже произведена)– расчет активностей плоскостей сложных характеристик.

Шаг.2 Выбор сложных клеток, активность которых превышает $U_{min}(C,v,\zeta)$: $\forall v: v \in Layer(C), U(C,v,\zeta) > U_{min}(C,v,\zeta)$. Нахождение областей ρ_i, включающих все выбранные сложные клетки и их начальные окрестности R.

Шаг.3 Для анализируемых ВНН, или *VTU*, – на первой итерации это все связанные ВНН – найти расстояние компонент вектора весов от выходов, ассоциированных с ними выбранных сложных клеток v:

$\forall VTU: VTU \in \Psi, d = \sqrt{\sum_{\rho \in M}\left(\delta \cdot w(\rho,\zeta) \cdot \left(w(\rho,\zeta) - U(C,\rho,\zeta)\right)\right)^2}$, где Ψ – область, анализируемых ВНН, $w(\rho,\zeta)$ – значение синаптического веса ВНН, связанного с клеткой ρ в плоскости ζ, $0 < \delta < 1$ – константа масштабирования

Шаг.4 Исключить из Ψ клетки с расстоянием d, превышающем D_{max}:

$$\Psi = \bigcup_j \mathrm{VTU}_j : d_j < D_{max}$$

Шаг.5 Если $\Psi = \emptyset$, то выдать сообщение об отсутствии такого образа в памяти, иначе шаг.6

Шаг.6 Если обработаны сигналы от всех сложных клеток $M \equiv \bigcup_i \rho_i$, то шаг. 7, иначе шаг.8.

Шаг.7 Выдать информацию о ВНН, входящих в Ψ, в порядке возрастания расстояния d_j

Следующий шаг зависит от направления анализа. Для восходящего анализа:

Шаг.8 Увеличить размеры регионов анализа (активные сложные клетки и их окрестности) на величину r. Если регионы пересекаются, то объединить их. На этом этапе происходит увеличение М. Перейти на шаг 3.

Для нисходящего анализа:

Шаг.8 Выбрать, еще не анализировавшиеся синапсы клетки с $w(\rho,\zeta) > w_{min}$ исключить из М области оценки восходящим методом и включить регионы ассоциированные с $w(\rho,\zeta)$ с новыми начальными окрестностями R (при самом первом запуске), если таковые синапсы отсутствуют увеличить R на r. Перейти на шаг.3.

Комбинированная стратегия одновременно использует шаг.8 для восходящего и нисходящего поиска.

Предложенный механизм ассоциативного поиска позволяет ускорить поиск решения как за счет корреляции близких максимально информативных фрагментов образа, так и за счет исследования наличия у образа существенных признаков.

Слой *классоассоциативных нейронов* (**КН**) служит для представления классов в ИНС. В процессе работы сети происходит усиление связей между ВНН и КН, если «учитель» указывает на существование класса образа. Усиление этой связи также происходит в дальнейшем, если система повторно распознает образ или ей принудительно укажут на этот фрагмент. Правило обучения:

$$u(\varsigma,j) = (1-\alpha)\cdot u(\varsigma,j) + \beta \cdot u(\varsigma,j)\cdot U^{-1}(VIEW,j)$$

где $u(\varsigma,j)$ – значение -того синаптического коэффициента КН ς, соединенного с ВНН j с максимальным выходным значением, α – скорость забывания, β – скорость обучения синапса, $U(VIEW,j)$ – выход ВНН.

Алгоритм обучения ИНС приведен на рис. 2.

Процесс распространения сигнала внутри иерархической ИНС можно представить следующими этапами:

1. Обнулить активности всех клеток сети.
2. Активировать клетки рецептивной плоскости выбранным сегментом, получить значения полутона как выходного сигнала для всех возбужденных клеток плоскости.
3. Рассчитать размеры рецептивных полей для нейронов уровня простых и подуровня простых локальных ориентаций слоя сложных характеристик.
4. Установить значения синаптических весов для S-нейронов.
5. Рассчитать активность S-клеток.
6. Группировать результаты всех S-клеток операцией MAX.
7. Произвести латеральное торможение в MAX-подуровне уровня простых характеристик.

8. Рассчитать активность С-клеток.
9. Произвести латеральное торможение MAX- (для подуровня сложных характеристик) и С- плоскостей.
10. Запуск ассоциативного поиска среди ВНН. Определить множество результатов – образов, наиболее похожих на входной сегмент.
11. Возбуждение классоассоциативных клеток.

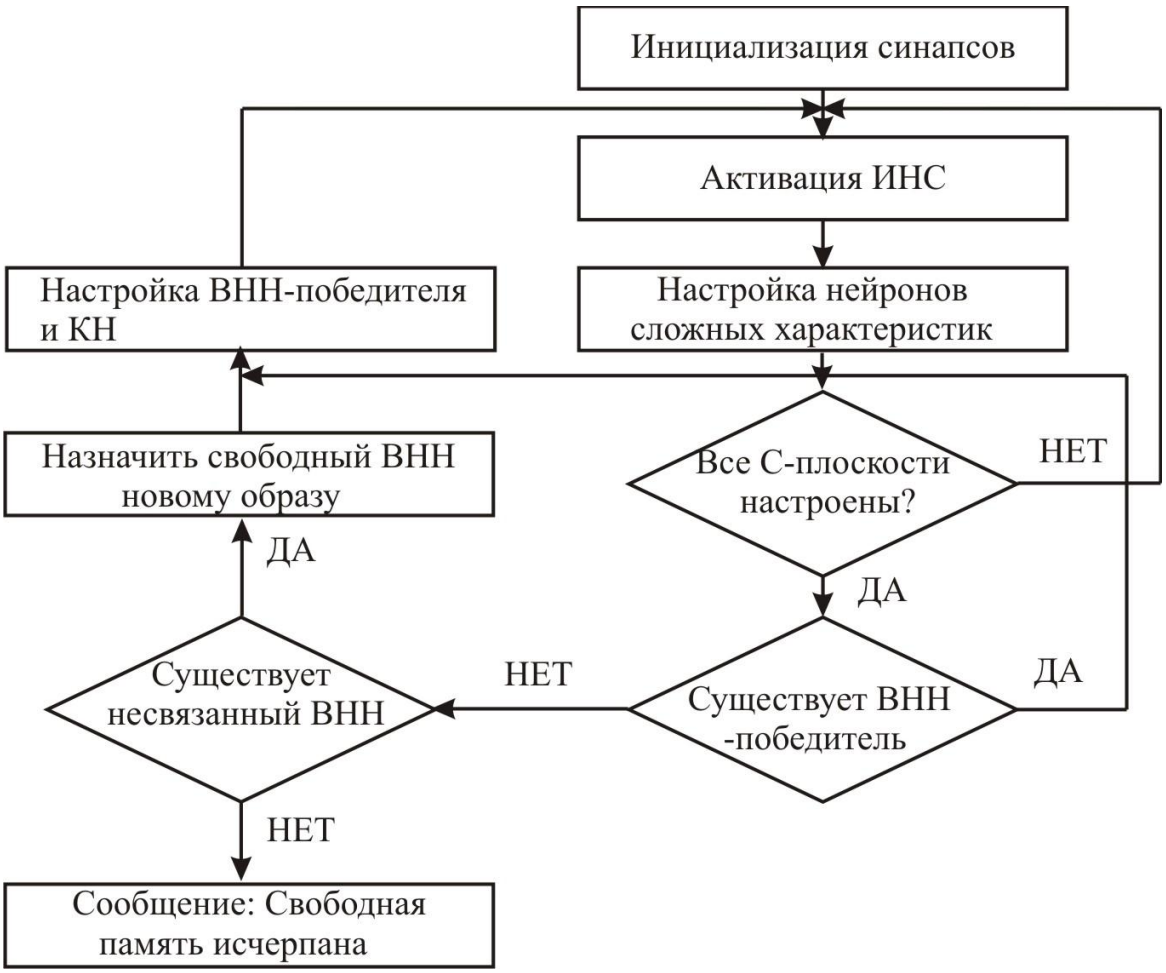


Рис. 2 Алгоритм обучения иерархической ИНС

В третьей главе дано описание системы анализа двумерных сцен на базе взаимодействия подсистем внимания и распознавания. Подсистема внимания включает в себя блок сегментации и блок восходящего внимания. Алгоритм восходящего внимания позволяет локализовать области сцены со значительным контрастом интенсивности, цветовых компонент и локальных ориентаций соседних областей сцены, полученных в нескольких масштабах (это делается посредством построения восьмиуровневой Гауссовой пирамиды. Для расчета также оцениваются предпочтительные ориентации $\theta \in \{0°, 45°, 90°, 135°\}$ разной ширины.

Результаты оценки каждого из трех каналов обработки проецируются на плоскости контраста. Контраст определяется как разность между активностью центра (клетки, или пикселя, на нижнем уровне пирамиды) и её окружением (клетки на верхнем уровне пирамиды):

$$\mathcal{I}(c,s) = |I(c) \ominus I(s)|,$$
$$\mathcal{RG}(c,s) = |(R(c) - G(c)) \ominus (R(s) - G(s))|,$$
$$\mathcal{BY}(c,s) = |(B(c) - Y(c)) \ominus (B(s) - Y(s))|,$$
$$\mathcal{O}(c,s) = |O(c) \ominus O(s)|,$$

где $\mathcal{I}(c,s)$, $\mathcal{RG}(c,s)$, $\mathcal{BY}(c,s)$ и $\mathcal{O}(c,s)$ значения контраста соответственно интенсивности, красной и зелёной компонент, синей и жёлтой компонент, локальных ориентаций между уровнями c и s Гауссовой пирамиды, $c \in \{2,3,4\}$, $s = c + \delta$, $\delta \in \{2,3\}$, $I(n)$, $R(n)$, $G(n)$, $B(n)$, $Y(n)$, $O(n)$ – значения компоненты карты (уровня) соответственно для интенсивности, красного, зелёного, синего, жёлтого оттенков и локальных ориентаций, $\ominus$ – разность центра и окружения. Активности карт контраста, соответствующих одной и той же переменной анализа, группируются для получения карт видимости соответственно для интенсивности, цветов и ориентаций:

$$\overline{\mathcal{I}} = \oplus_{c=2}^{4} \oplus_{s=c+3}^{c+4} [N(\mathcal{I}(c,s))],$$
$$\overline{\mathcal{C}} = \oplus_{c=2}^{4} \oplus_{s=c+3}^{c+4} [N(\mathcal{RG}(c,s)) + N(\mathcal{BY}(c,s))],$$
$$\overline{\mathcal{O}} = \sum_{\theta \in \{0°,45°,90°,135°\}} N\left(\oplus_{c=2}^{4} \oplus_{s=c+3}^{c+4} [\mathcal{O}(c,s,\theta)]\right),$$

где $N(\dots)$ – нормализация, $\oplus$ – сложение активностей карт разных масштабов.

После нахождения активностей всех плоскостей видимости выходы последних комбинируются в одном двумерном массиве клеток, называемом картой особенностей:

$S = f(N(\overline{\mathcal{I}}), N(\overline{\mathcal{C}}), N(\overline{\mathcal{O}}))$, где f – функция, аргументы которой указывают на вклад каждого канала и настраиваются с учителем на основе обучающей выборки. Расположение наиболее активных клеток карты особенностей, найденное согласно процедуре WTA, позволяет локализовать наиболее информативные области. Другими словами, положение наиболее активной клетки указывает на фрагмент сцены, который попадет в высшие области (центральное представление) для последующей обработки в первую очередь. Далее, выделяются области сцены, ассоциированные со вторым по активности нейроном карты особенностей и т. д. Таким образом, фокус внимания будет переходить от наиболее контрастных сегментов сцены к более однородным.

Практически для локализации регионов внимания выбирается значение минимальной величины значимой особенности S_{min} и в области внимания $A_k : v \in A_k, v \in S, S(v) \geq S_{min}$. Порядок попадания в фокус внимания определяется либо максимальным значением активности каждого региона $\exists v : \forall z\, v, z \in A_k, v, z \in S, z \neq v, S(v) \geq S(z)$, либо средним значением активности для всей области $S_{avg} = \frac{1}{n} \sum_{i=1}^{n} S(v_i), \forall v_i \in A_k$, n – мощность A_k.

Для реализации блока сегментации был взят за основу метод Монте-Карло на базе сети Маркова, со стимуляцией восходящими процессами анализа данных. При использовании алгоритма восходящего анализа сегментацию можно обозначить вектором скрытых переменных W, которые описывают состояние для генерации сцены I: $W = (K, \{(R_i, l_i, \Theta_i); i = 1,2, \dots, K\})$, где Θ_i

представляет собой стохастический процесс, l_i – модель региона (анализируемые семейства представлены на рис.3), K – количество связанных сегментов. Каждый регион сцены I_R обязан быть когерентным, в том смысле, что I_R есть реализация вероятностной модели $p(I_R, \Theta)$. Согласно процедуре обработки в статистической среде Байеса можно произвести вывод вектора W из I на пространстве всех решений Ω:

$$W \sim p(W|I) \propto p(I|W)p(W), \qquad W \in \Omega$$

Однородная Зашумление Текстура Оттенивание

Рис.3. Семейства образов сцены

Процедура анализа заключается в следующем: сначала осуществляется предварительная оценка сцены известными методами исследования градиента интенсивности и кластеризации оттенков цветов. Параметры этого решения будут являться одним из состояний цепи Маркова. Затем путем случайных вариаций производится поиск более оптимальных решений. Расчет степени генерации сцены при известной сегментации $p(I|W)$ зависит от того насколько каждый сегмент соответствует одной из моделей, связанных сегментов естественных сцен:

$$p(I|W) = \prod_{i=1}^{K} p(I_{R_i}; \Theta_i, l_i)$$

При функционировании процедуры стохастического поиска предполагается получение множественного решения, присущего неоднозначности понятия сегментации сцен. Пример такой серии приведен на рис.4.

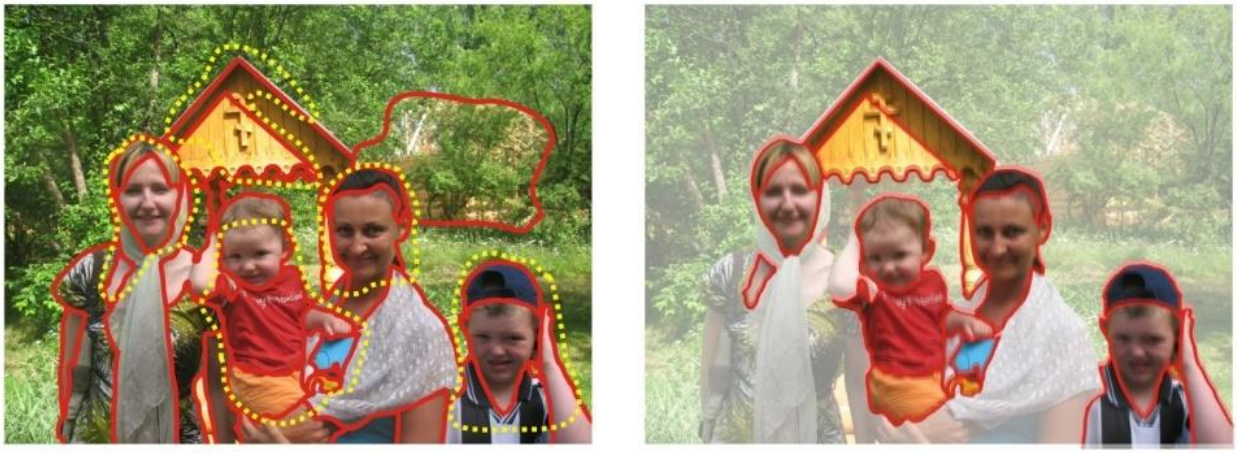
P(I|W)=0,723 *P(I|W)=0,742* *P(I|W)=0,739*

Рис.4 Пример серии решения для одной сцены

Для решения задачи выделения контрастных сегментов был разработан алгоритм интеграции блоков внимания для получения решения в виде множества наиболее значимых сегментов, (которые также могут пересекаться!).

Алгоритм интеграции предполагает первоначальное получение регионов $A_k: v \in A_k, v \in S, S(v) \geq S_{min}$ и серии N сегментаций обозначенной T: $\forall W_i: W_i \in T, P(W|I) \geq P(W|I)_{min}$, где $P(W|I)_{min}$ показатель, названный минимальной степенью адекватности. Далее определяется степень наложения регионов W_i и областей A_k. Для всех областей R_{ij} сегментаций W_i, пересекающихся с A_k находится:

$$P(R_{ij}|A_k) = {S(R_{ij} \cap A_k)}/{S(R_{ij})}$$



Регионы фокуса внимания пример сегментации с $P(I|W)=0,615$

Регионы сегментации с $P(R_{ij}|A_k)>0,7$

Рис.5 Результат интеграции восходящего внимания и сегментации

Дальнейшие расчеты требуют задания значения P_{min} минимальной вероятности интереса. Все регионы со значением $P(R_{ij}|A_k)$ превышающим P_{min} являются «интересными» для последующей обработки, а все оставшиеся отбрасываются. На заключительной стадии происходит возбуждение иерархической ИНС прошедшими отбор фрагментами R_{ij} с вероятностями $P(R_{ij}|A_k)$.

На рис.5 в левой части приведен пример одной из вероятных сегментаций (регионы разделены непрерывающимися кривыми) и регионов внимания, границы последних прерывистые. В правой части того же рисунка – отмечены регионы, активирующие иерархическую ИНС и полученные в результате работы описанного алгоритма.

Четвертая глава посвящена практической реализации и тестированию системы. На рис.6 показана главная форма разработанного приложения с открытыми дочерними окнами «Структура ИНС», «Визуализация значимых контрастных регионов» и «Результаты распознавания».

Программа выполняет следующие функции:

- Предоставление возможности построения структуры ИНС с использованием конструктора или с помощью мастеров.
- Формирование и настройка выборки для обучения подсистем внимания и распознавания.
- Визуализация процессов активации подсистем ИНС.

- Обучение модели алгоритмами, заданными пользователем.
- Интерпретация результатов возбуждения ИНС.

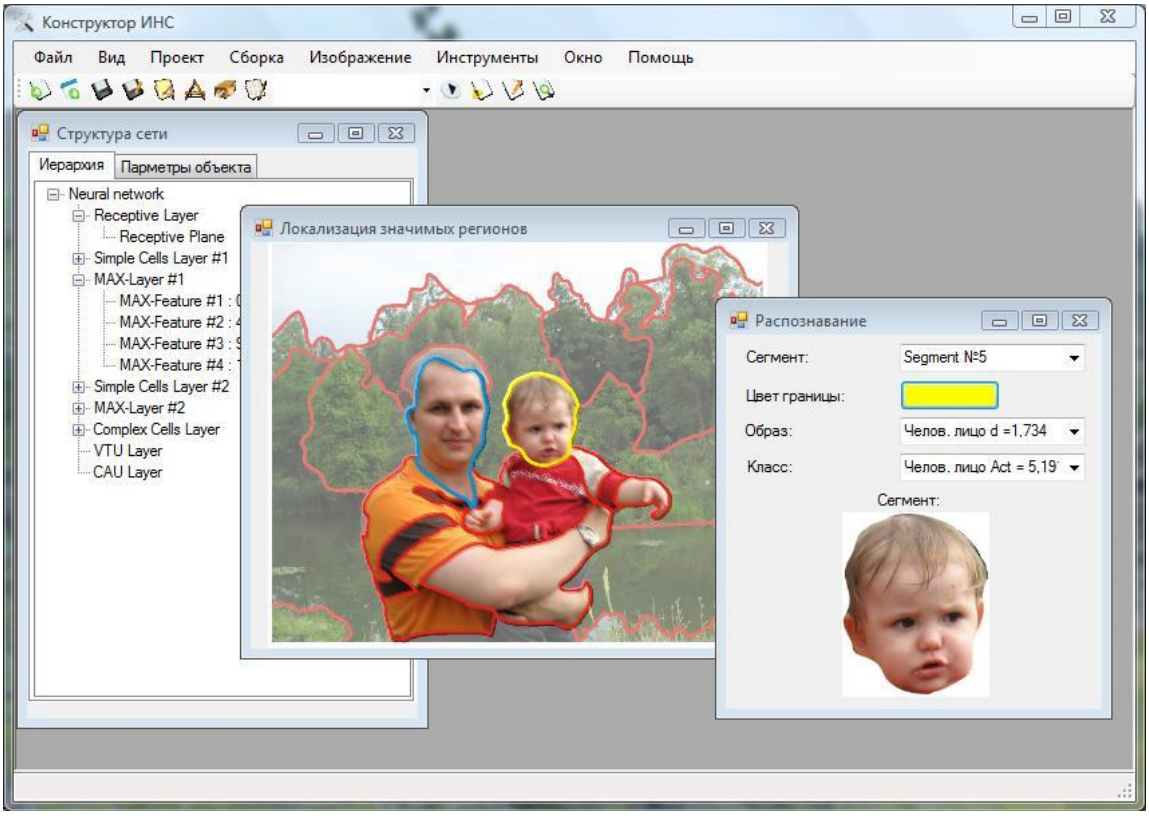


Рис.6 Вид программного комплекса

Возможности разработанной математической модели и программного комплекса проверялись путем тестирования результатов работы модели с результатами моделей Ризенхубера, Вальтера и ЛеКуна. Для сравнения особенностей каждой модели для решения задач интерпретации изображений приведены в таблице 2.

Табл.2 Характеристики моделей ИНС

Модель	Обучение	Включение новых образов	Внимание	Пластичность весов (изменчивость знаний ИНС)
Иерархическая ИНС	Сочетание алгоритмов обучения с учителем и без учителя	Требует только включения ВНН для каждого нового образа, обучение в процессе функционирования ИНС	Восходящее внимание	Как один из базовых принципов
Вальтера (расширение HMAX Ризенхубера)	Отсутствует (веса предустановлены)	Требует только включения настроенного RBF-нейрона для каждого нового образа	Восходящее внимание	Невозможно
Расширенная LeNET ЛеКуна	Классические алгоритмы оптимизации	Требует полного переобучения сети	Отсутствует	После обучения, ресурсоемко

Тестирование осуществлялось для двух типов задач. Первый вариант «классический» предполагает, что на вход модели поступают фрагменты сцены, содержащие нормированный и отцентрированный образ человеческого лица, т.е. влияние фона и посторонних образов минимально, объект представлен в нужном масштабе (Такая задача свойственна системам персональной идентификации). Для того чтобы показать надежность модели требовалось оценить отклики сетей при двух видах повреждения входного вектора: при зашумлении изображения случайными пикселями и при нарушении прямоугольного фрагмента сцены. Результаты по первому тесту приведены в таблице 3.

Табл.3. Процент числа корректных ответов первого теста на выборке из 250 изображений,%

Модель	Без шумов	10% случайных шумов	25% случайных шумов	Нарушенный фрагмент 10% площади	Нарушенный фрагмент 25% площади
Иерархическая ИНС	81,6	79,7	75,8	78,1	74,8
Модель Ризенхубера	83,4	81,3	74,2	67,4	45,3
Расширенная модель ЛеКуна	78,5	70,1	61,5	66,3	52,7

Видно, что возможности предлагаемой модели по сравнению с другими оцениваемыми для не зашумленного изображения отличаются незначительно, однако если оценивать надежность распознавания для поврежденной сцены, заметно главное достоинство предлагаемой сети – концентрации на существенных фрагментах образа.

Второй тест оценивает возможности моделей при анализе естественных сцен, т.е. задача усложняется за счет необходимой локализации образов. Задача решалась для изображений, содержащих 3, 5 и 7 человеческих лиц. В таблице 4 даны результаты по каждой серии изображений.

Результат предлагаемой модели обусловлен взаимодействием блоков внимания и распознавания. Проведенные эксперименты позволяют говорить о перспективности разработанного подхода и о возможности его использования в системах анализа визуальных данных. При возможных запросах со стороны потребителей разработанный программный код при дополнительной доработке может выступать как ядро системы интеллектуальной визуальной обработки промышленных данных.

Табл.4 Результаты правильности локализации и распознавания лиц моделей на втором тесте (выборка 150 изображений), %

Модель	3 фрагмента лиц	5 фрагментов лиц	7 фрагментов лиц
Иерархическая ИНС	75,2	78,2	72,8
Модель Вальтера	56,7	49,3	37,6
Модель Вальтера с дополнительным анализом цветовых оттенков кожи	78,6	76,4	72,1

3. ЗАКЛЮЧЕНИЕ

Исследования, проведенные в диссертационной работе, были направлены на решение актуальной задачи распознавания образов, расположенных на произвольной сцене. Математическая модель интеграции процессов внимания и распознавания представляется основой для создания промышленных и исследовательских систем интерпретации двумерных сцен.

1. На основании экспериментов и процедуры создания системы, проведенных в диссертации, показано, что предлагаемый подход не только обеспечивает работу со сценами любой сложности, но и обладает возможностями адаптации и расширения в соответствии с запросами пользователя.
2. Разработанная модель обладает более эффективным алгоритмом интерпретации образов, расположенных на сценах, чем существующие современные модели, основанные на встречном распространении сигналов. Это является следствием, фокусировки ресурсов на значимых регионах сцены и ассоциативного характера поиска в пространстве запомненных образов.
3. Алгоритм интеграции процессов восходящего внимания и сегментации, позволяет, во-первых, сократить размерность обрабатываемых подсистемой распознавания регионов сцены за счет учета контраста визуальных параметров и, во-вторых, подчеркнуть возможность получения серии интерпретаций образов для одних и тех же областей.
4. Разработанный программный комплекс позволяет решить задачу интерпретации образов на визуальных сценах. Помимо исследованной практической задачи комплекс обладает достаточным инструментом для создания ИНС любой структуры и набором классических алгоритмов настройки, как с учителем, так и без.
5. Предложенная математическая модель, алгоритмы обучения и функционирования, а также программный комплекс использованы в учебном процессе на кафедре Оптимизации систем управления ТПУ, на кафедре Программирования ТГУ, внедрены в Институте оптики атмосферы СО РАН (г.Томск).

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Аксёнов С.В., Новосельцев В.Б. Организация и использование нейронных сетей. Томск. Изд-во НТЛ, 2006 – 128 стр.

2. Аксёнов С.В., Новосельцев В.Б. Повышение качества распознавания сцен нейронной сетью неокогнитрон // Известия Томского политехнического университета. – 2006. Т.309, №7.– С.87–91.
3. Аксёнов С.В. Механизмы выделения контрастных фрагментов сцены системой восходящего внимания // Известия Томского политехнического университета. – 2008. Т.312, №5. – С.97-101.
4. Аксёнов С.В., Танцов Д.К. Система видеоконтроля транспортируемых бревен // «Научная сессия ТУСУР–2004»: труды Всероссийской научно-технической Конференции / ТУСУР, – Томск, 2004. Т.1. – С. 25–28.
5. Аксёнов С.В. Модификации алгоритма обучения персептрона методом обратного распространения // «Научная сессия ТУСУР–2004»: труды Всероссийской научно-технической конференции / ТУСУР, – Томск, 2004. Т.2. – С.42-44.
6. Аксёнов С.В. Использование нейронной сети ART-2 при классификации пациентов с осложнениями инфаркта миокарда // «Молодежь и современные информационные технологии – 2006»: труды III Всероссийской научно-технической конференции / ТПУ, – Томск, 2006. – С.174–175.
7. Аксёнов С.В., Новосельцев В.Б. Построение самоорганизующейся иерархической нейронной системы накопления признаков // «Молодежь и современные информационные технологии – 2007»: труды V Всероссийской научно-технической конференции / ТПУ, – Томск, 2007. С.217–219.
8. Аксёнов С.В. Самоорганизующаяся модель иерархической нейронной сети для распознавания сцен // «Инновационные недра Кузбасса. IT-технологии»: труды VI Всероссийской научно-технической конференции / КГТУ, – Кемерово, 2007. С.132–133.
9. Аксёнов С.В. Взаимодействие систем внимания и распознавания для интерпретации сцен // «Современные техника и технологии 2007» : труды XIII Международной научно-практической конференции студентов и молодых ученых / ТПУ, – Томск, 2007. Т.2, С.268–270.
10. Аксёнов С.В. Самоорганизующееся построение иерархической нейронной сети для анализа сцен // «Студент и научно-технический прогресс»: труды XLV Международной научно-технической конференции. / НГТУ, – Новосибирск, 2007. Т.4. – С.156–157.
11. Аксёнов С.В. Взаимодействие систем распознавания и внимания для интерпретации сцен // «Молодежь и современные информационные технологии – 2008»: труды VI Всероссийской научно-технической конференции / ТПУ, – Томск, 2008. – С.319–321.
12. Аксёнов С.В. Алгоритм рекуррентного взаимодействия в системе анализа зрительной информации на базе иерархических нейронных сетей // «Инфокоммуникационные технологии в науке, производстве и образовании»: труды II Международной научно-технической конференции / СевКавГТУ, – Кисловодск, 2008. Т.2. – С.56–57.

13. Аксёнов С.В. Иерархическая нейронная система для анализа объектов, расположенных на произвольном фоне // // «Студент и научно-технический прогресс»: труды XLVI Международной научно-технической конференции / НГТУ, – Новосибирск, 2008. Т.4. – С.112–113.
14. Аксёнов С.В. Методы иерархического анализа многообразных сцен // «Информационные технологии в профессиональной деятельности и научной работе»: труды Всероссийской научно-технической конференции/ МарГТУ, – Йошкар-Ола, 2008. Т.2. – С.9–12.