

ИНВАРИАНТНОЕ ПРЕДСТАВЛЕНИЕ И РАСПОЗНАВАНИЕ ОБЪЕКТОВ С ПОМОЩЬЮ ОКНА "ВНИМАНИЯ" И СПЕЦИФИЧЕСКОЙ ТРАЕКТОРИИ ОСМОТРА

*В.И.Гусакова, И.А.Рыбак Л.Н.Подладчикова,
А.В.Головань, Н.А.Шевцова*

Рассматривается модель системы для инвариантного представления и распознавания объектов на полутоновых изображениях. Система выполняет анализ изображений на нескольких уровнях разрешения и формирует инвариантные признаки фрагментов и объектов в целом. Признаки первого рода — 49-мерные векторы значений входных сигналов — формируются при каждой фиксации окна "внимания" в результате параллельной обработки фрагментов изображения в базовой (центральной) и контекстных (периферических) точках. Признаки второго рода — траектории фиксаций окна "внимания", характерные для каждого объекта — формируются последовательно в результате осмотра изображения.

Описаны алгоритмы, позволяющие формировать траекторию осмотра автоматически. Возможная базовая точка для каждой следующей фиксации окна "внимания" выбиралась из 48 контекстных точек предыдущей фиксации в зависимости от ряда признаков фрагмента изображения.

В компьютерных экспериментах обнаружена взаимосвязь между некоторыми параметрами и определены условия, при которых имеет место сходство траекторий, формируемых с помощью супервизорного и автоматического алгоритмов. Разработанные алгоритмы обеспечивают распознавание объектов инвариантно к масштабу, сдвигу, повороту и зашумлению. По результатам машинных экспериментов сделан ряд предсказаний о возможных свойствах реальной зрительной системы. Коротко обсуждаются возможные пути аппаратной реализации системы для инвариантного представления и распознавания объектов.

1. ВВЕДЕНИЕ

При разработке систем технического зрения, в частности, ориентированных на анализ сложных полутоновых изображений, существует множество проблем представления и распознавания зрительной информации, не нашедших еще своего эффективного решения. Многие из этих проблем, с одной стороны, связаны с большим количеством информации, которая должна быть обработана в реальном времени (объем и скорость вычислений), а, с другой стороны, с искажениями образов в реальном мире (эффективность и инвариантность распознавания).

Можно ожидать, что проблемы первого рода будут решены в перспективе с помощью усовершенствования стандартных технических средств, в то время как проблемы второго рода будут более эффективно решены на основе воспроизведения механизмов зрительного восприятия.

В настоящее время в этой области начата разработка систем нового класса — фовеальных систем [1–4], которые имитируют два взаимосвязанных свойства зрительной системы человека: неоднородность плотности рецепторов сетчатки и глазные движения. Эти свойства обеспечивают значительное уменьшение количества параллельно обрабатываемой информации и фиксацию фовеа (части сетчатки с наибольшей плотностью рецепторов) на фрагментах изображения, которые должны быть обработаны с высоким пространственным разрешением. Один из вариантов фовеальной системы, MARR, в котором был реализован супервизорный режим формирования траектории осмотра объектов на этапе запоминания (обучения) был разработан нами ранее [3]. В этом режиме точки фиксации окна "внимания" на изображении объекта выбирались оператором. Вместе с тем, для увеличения эффективности обработки зрительной информации с помощью системы MARR необходимы разработка алгоритмов автоматического (несупервизорного) формирования траекторий осмотра объектов и поиск критериев выбора точек фиксации. Исследование этих проблем и было задачей данной работы.

1. СВОЙСТВА СИСТЕМЫ MARR

Основными свойствами системы MARR, обеспечивающей инвариантное представление и распознавание зрительных объектов являются следующие [3]:

1. Существенное сжатие входной информации за счет специфической структуры окна "внимания" (рис.1).

Представление и обработка информации осуществляется в 49 локальных точках фрагмента изображения — базовой и 48 контекстных (на пересечении 8 секущих и 3 окружностей) точках — при каждой фиксации окна "внимания".

2. Обработка информации на нескольких уровнях разрешения.

Структура окна "внимания" обеспечивает обработку зрительной информации в базовой и контекстных точках на уровнях разрешения, различающихся на фиксированную величину (см.рис.1). В базовой точке информация обрабатывается с наибольшим уровнем разрешения, в контекстных точках уровень разрешения экспоненциально убывает с удалением от центра (рис.2). Кроме того, уровень разрешения в базовой точке может быть выбран в зависимости от размера элементов данного фрагмента изображения.

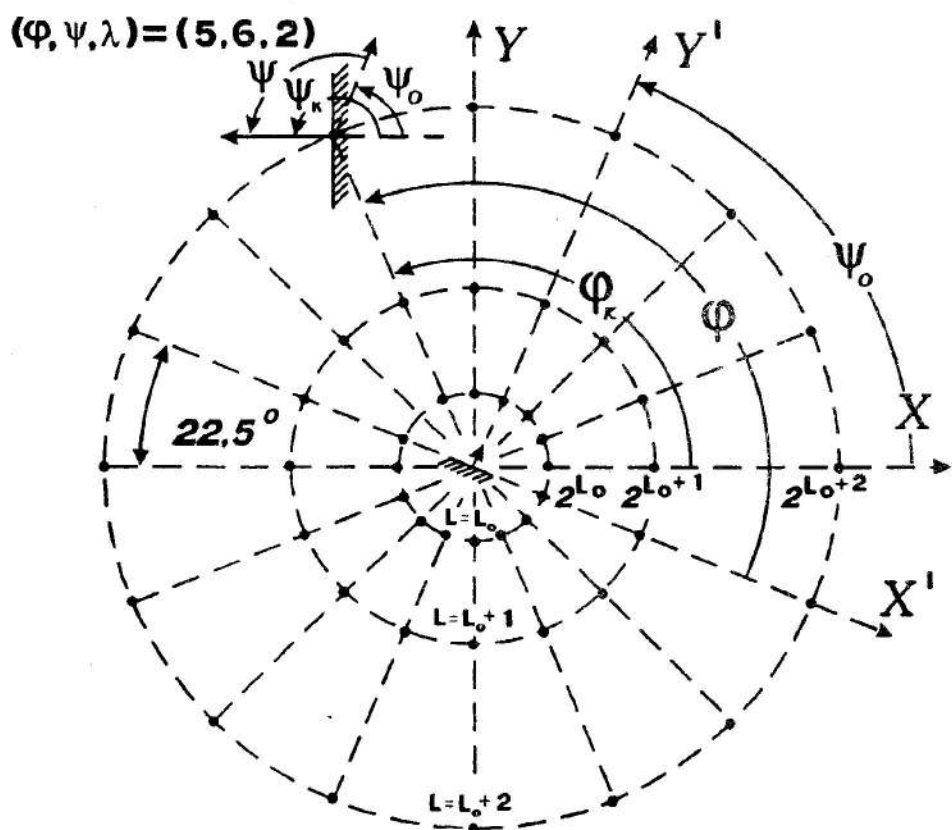


Рис. 1.

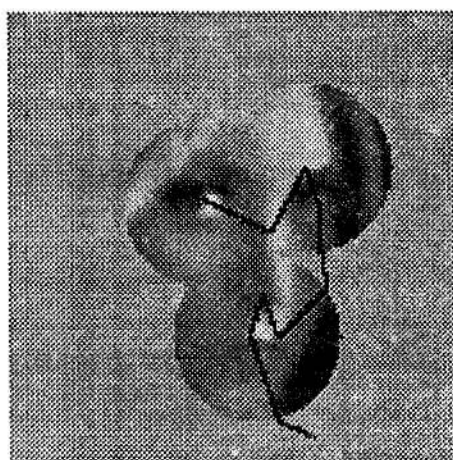


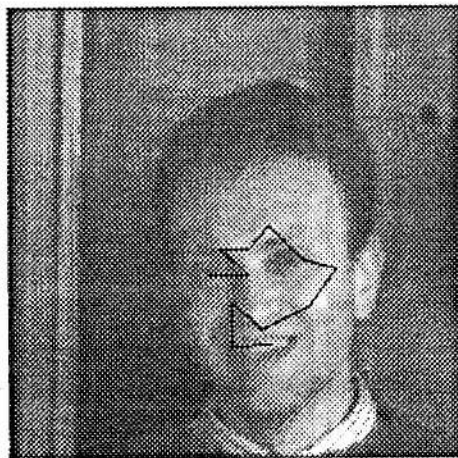
Рис. 2.

3. Выделение ориентированных линейных сегментов в окрестностях базовой и контекстных точек окна "внимания".

Использование ориентационной фильтрации, характерной как для многих систем технического зрения, так и для зрительной коры высших животных [5-7], обеспечивает значительное сжатие информации, обрабатываемой и хранящейся в памяти.

4. Специфичность траектории последовательного осмотра изображений (рис.3) и сходство траекторий осмотра данного объекта на этапах запоминания (обучения) и распознавания.

Это свойство системы MARR основано на психофизических данных о параметрах траектории глазных движений в процессе зрительного восприятия человека [8, 9].



Перечисленные свойства системы MARR обеспечивают существенное сжатие обрабатываемой информации и позволяют формировать инвариантные признаки фрагментов и объектов в целом. Инвариантные признаки фрагментов представлены 49-мерными векторами величин входных сигналов, которые формируются в результате параллельной обработки информации в базовой и

контекстных точках при каждой фиксации окна "внимания". Инвариантные признаки объектов в целом (специфические траектории осмотра) формируются последовательно в процессе сканирования объекта на этапах запоминания и распознавания.

2. АЛГОРИТМЫ АВТОМАТИЧЕСКОГО ФОРМИРОВАНИЯ ТРАЕКТОРИИ ОСМОТРА ОБЪЕКТОВ

При разработке алгоритмов автоматического формирования траектории осмотра изображения мы исходили из ряда условий и допущений. Так, каждая следующая точка траектории осмотра выбирается из 48 контекстных точек, определенных с помощью окна "внимания" в данной точке фиксации. В качестве базовой точки для следующей фиксации среди контекстных точек желательно выбрать ту, ориентированные сегменты контрастных границ в окрестностях которой имеют максимальную яркость, максимальную величину контраста и представлены в наибольшем количестве. Желательно, чтобы эта точка располагалась на большем расстоянии от базовой точки, т.е. на более низком уровне разрешения, и ориентация сегментов в ней максимально отличалась от таковой в базовой точке. Очевидно, выбираемая в качестве базовой точка не должна совпадать с уже полученными точками траектории и располагаться близко от границы изображения.

В качестве показателя яркости сигналов в точке (x, y) изображения будем брать среднюю яркость в ее окрестности — $I_m(x, y)$

$$I_m(x, y) = \frac{1}{N_1} \sum_{(x_i, y_i) \in \Omega_1(x, y)} I(x_i, y_i), \quad (1)$$

где $I(x_i, y_i)$ — яркость сигналов в точке (x_i, y_i) ; $\Omega_1(x, y)$ — окрестность точки (x, y) ; N_1 — количество точек, попавших в эту окрестность.

Величину контраста линейного сегмента $E_m(x, y)$ в точке (x, y) определим соотношением:

$$E_m(x, y) = \frac{1}{N_2} \sum_{(x_i, y_i) \in \Omega_2(x, y)} E(x_i, y_i), \quad (2)$$

где $E(x_i, y_i)$ — величина контраста линейного сегмента изображения в точке (x_i, y_i) , $\Omega_2(x, y)$ — окрестность точки с (x, y) ; N_2 — количество точек, попавших в эту окрестность.

В качестве показателя количества контрастных элементов будем брать число выделенных ориентированных сегментов в окрестности точки (x, y)

$$O(x, y) = \sum_{(x_i, y_i) \in \Omega(x, y)} O(x_i, y_i), \quad (3)$$

где $\Omega(x, y)$ — контекстная область для точки (x, y) , а

$$O(x_i, y_i) = \begin{cases} 1 & \text{если } 0 \leq \varphi(x_i, y_i) \leq 15; \\ 0 & \text{в противном случае.} \end{cases}$$

$\varphi(x_i, y_i)$ — ориентация сегмента в области с координатами центра (x, y) .

Различие ориентаций сегментов в базовой и контекстных областях определяется показателем $D(x, y)$

$$D(x, y) = |\varphi(x_b, y_b) - \varphi(x, y)|, \quad (4)$$

а уровень разрешения показателем — L .

Опишем рекуррентный процесс формирования последовательности точек траектории осмотра изображения $\{(x_b, y_b), (x^t, y^t) \dots\}$.

Пусть (x^t, y^t) — текущая точка траектории. (В качестве начальной точки (x^0, y^0) можно взять центр изображения или точку, указанную оператором).

Для того, чтобы задать приоритет точек (x, y) , которые ранее не принадлежали окрестности $\Omega_3(x^t, y^t)$ точек (x^t, y^t) траектории, введем функцию

$$P(x, y, t) = \begin{cases} U_{xy}(0) & = I(x, y), \\ U_{xy}(t-1) - \alpha & \text{для } (x, y) \in \Omega_3; \\ U_{xy}(t-1) + \beta & \text{в противном случае,} \end{cases} \quad (5)$$

где α, β — константы, определяемые экспериментально. Тогда в качестве следующей точки траектории возьмем точку (x^{t+1}, y^{t+1}) , дающую максимум функции

$$F(x, y) = k_1 \cdot I_m(x, y) + k_2 \cdot E_m(x, y) + k_3 \cdot O(x, y) + \\ + k_4 \cdot D(x, y) + k_5 \cdot P(x, y, t) + k_6 \cdot L, \quad (6)$$

где $(x, y) \in \Omega(x^t, y^t)$.

В описанных выше алгоритмах использовались абсолютные величины яркости и контраста. Был разработан также алгоритм формирования траекторий осмотра на основе их относительных величин. В этом случае относительная яркость для точке (x, y) определялась следующим образом:

$$I_m(x, y) = \frac{I(x, y)}{1/N^2 \sum_{i=1}^N \sum_{j=1}^N I(x_i, y_j)}, \quad (7)$$

а величина контраста:

$$E_m(x, y) = \frac{E(x, y)}{1/N^2 \sum_{i=1}^N \sum_{j=1}^N E(x_i, y_j)}, \quad (8)$$

где N^2 — количество пикселей изображения. Потенциальная функция $P(x, y, t)$ определялась следующим образом:

$$P(x, y, t) = \begin{cases} U_{xy}(0) & = I_m(x, y), \\ U_{xy}(t-1) - \alpha & \text{для } (x, y) \in \Omega(x, y), \\ U_{xy}(t-1) + \beta & \text{в противном случае,} \end{cases} \quad (9)$$

где α_m, β_m — константы, определяемые в соответствии со средней яркостью изображения. Критерий выбора следующей точки фиксации был тем же, что и ранее, т.е. для того, чтобы одна из контекстных точек окна "внимания" была выбрана в качестве следующей базовой, в ней должен достигаться максимум функции $F(x, y)$ (6).

Разработанные алгоритмы автоматического формирования траекторий осмотра изображения объекта с помощью фиксации окна "внимания" были тестированы в машинных экспериментах при обработке тестовых изображений. Эти эксперименты показали удовлетворительное функционирование алгоритмов и формирование траекторий, обеспечивающих распознавание объектов инвариантно к масштабу, сдвигу, повороту и зашумлению. Определены также величины параметров, при которых имеет сходство траекторий осмотра, формируемых с помощью supervisory и автоматического алгоритмов (рис.4). Кроме того, в машинных экспериментах было обнаружено, что можно упростить выражение $F(x, y)$ (6) положив равным нулю ряд коэффициентов.



Рис. 4а, б.



3. ВОЗМОЖНАЯ АППАРАТНАЯ РЕАЛИЗАЦИЯ СИСТЕМЫ MARR

Блок-схема разработанной системы для инвариантного представления и распознавания зрительных изображений MARR приведена на рис.5.

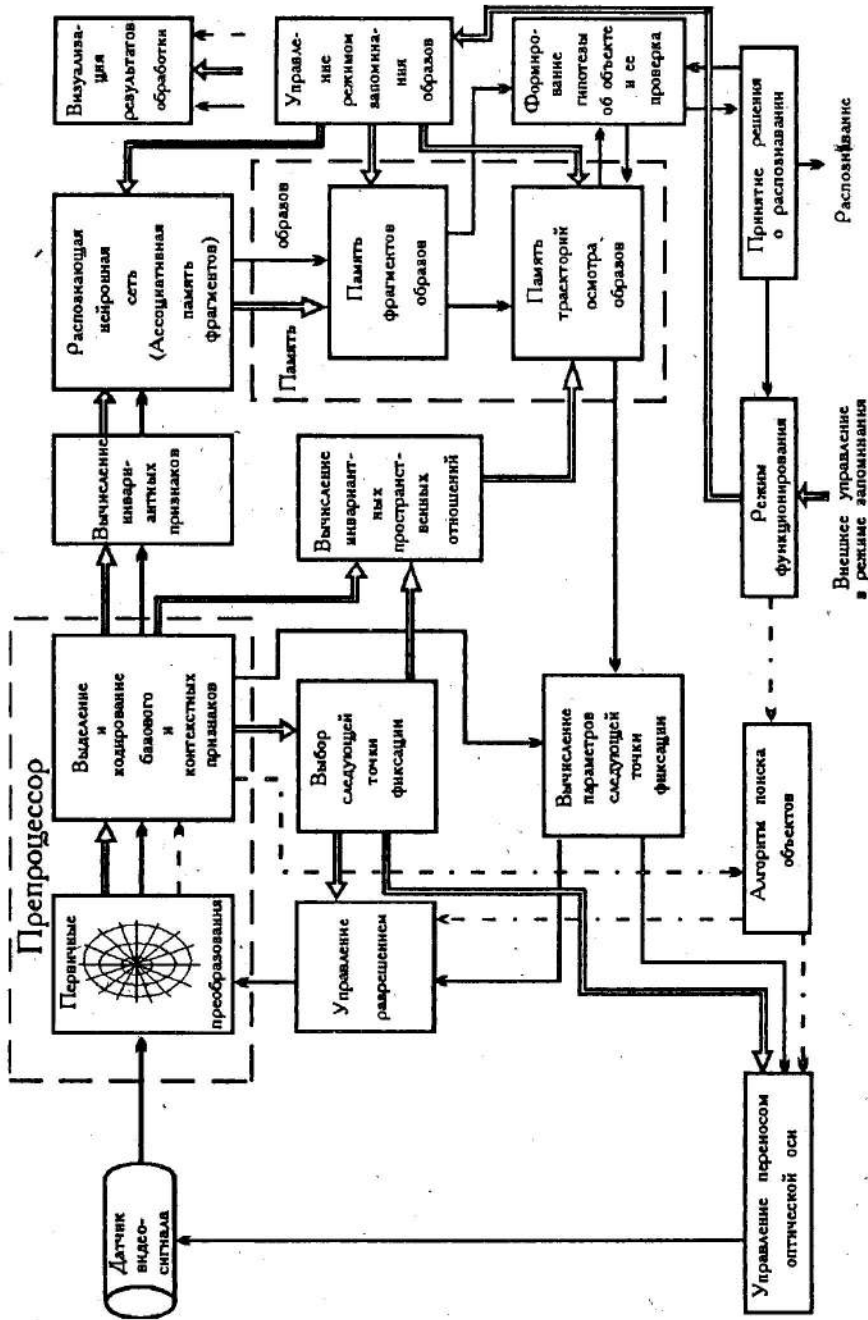


Рис. 5 Блок-схема системы MARR

В настоящее время все блоки этой системы программно реализованы на IBM PC AT. Возможны несколько вариантов аппаратной реализации блоков системы MARR, которые позволят существенно ускорить обработку информации, в частности, с помощью специализированных плат, вы-

полняющих ориентационную фильтрацию и распараллеливающих процедуры вычислений в базовой и контекстных точках. Одним из наиболее интересных блоков для аппаратной реализации, в том числе с помощью оптических технологий, является окно "внимания".

Один из вариантов оптико-механической реализации окна "внимания" предложен в [4].

4. ПРЕДСКАЗАНИЯ

В настоящее время известно очень мало о механизмах выбора последовательных точек фиксации в реальной зрительной системе [1, 8]. Результаты проведенных машинных экспериментов позволили сделать ряд предсказаний о возможных свойствах биологического прототипа. В плане правомерности таких предсказаний следует отметить два момента. Во-первых, разработанная модель системы MARR имитирует ряд свойств реальной зрительной системы, во-вторых, специфическую траекторию осмотра данного изображения в супервизорном режиме (взятую как эталонную при подборе параметров в автоматическом режиме) формировал оператор в соответствии собственными механизмами зрительного восприятия. Приведем некоторые из этих предсказаний:

1. наличие пространственно дискретных центров обработки зрительной информации в поле зрения (в ретинокорковой системе), которые могут быть центрами притяжения при выборе следующей точки фиксации (ранее [10] получены факты на уровне рецептивных полей нейронов зрительной коры, согласующиеся с этим предсказанием);
2. наличие критической разности между ориентацией линейных сегментов, расположенных в фовеальном и периферическом зрении, которая определяет переключение взгляда на данный фрагмент изображения;
3. углы и пересечения ориентированных сегментов зрительных объектов являются признаками, определяющими перевод взгляда.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант N 94-04-20803.

ЛИТЕРАТУРА

1. Burt P.J. // IEEE. 1988. V.76. P.1006.
2. Zeevi Y.Y and Ginosar R. Neural Computers for Foveating Vision Systems. / R.Eckmiller. — Advanced Neural Computers. Elsevier Science Publishers B.V.: North-Holland, 1990. P.323.

3. Rybak I.A., Golovan A.V., Gusakova V.I., Shevtsova N.A. and Podladchikova L.N. // *Neural Network World*, 4. 1991. P.245.
4. Hecht-Nielsen R. and Zhou Y.T. // *Report on Government Conference on Neural Networks*. — Ohio: USA, 1992.
5. Hubel D.H. and Wiesel T.N. // *Journal of Physiology*. 1962. V.160. P.106.
6. Marr D. *Vision*. / W.H.Freeman and Company Press, New York, 1982.
7. Rybak I.A., Shevtsova N.A., Podladchikova L.N., and Golovan A.V. // *Neural Networks*. 1991. V.4. P.3.
8. Yarbus A.L. *Eye movement and vision*. — Plenum Press: New York, 1967.
9. Noton D. and Stark L. // *Science*. 1971. V.171. P.72.
10. Rybak I.A., Podladchikova L.N., Shevtsova N.A., Golovan A.V., Gusakova V.I., Sandler V.M. *Proc.RNNS*. // *IEEE Symposium on Neuroinformatics and Neurocomputers*. Rostov-on-Don, Russia. 1992. V.1. P.323.

НИИ нейрокибернетики
им.А.Б.Когана РГУ,
Ростов-на-Дону,
Пеннсильванский университет,
Филадельфия, США

Поступила в редакцию
16 сентября 1993 г.

**INVARIANT REPRESENTATION AND RECOGNITION OF OBJECTS
BY MEANS OF "ATTENTION" WINDOW
AND SPECIFIC VIEWING TRAJECTORY**

*V.I.Gusakova, I.A.Rybak, L.N.Podladchikova,
A.V.Golovan, N.A.Shevtsova*

A model of a system for invariant representation and recognition of objects on grey-level images is considered. The system performs image analysis at several resolution levels and forms invariant features of fragments (the vectors of values of input signals in each fixation point of "attention" window) as well as of the object as a whole (specific trajectories of fixation points characteristic for the given object).