

# ИНФОРМАТИКА И УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМАХ

УДК 004.931

**В. Е. Гай, В. А. Утробин, Е. Н. Викулова,  
Н. А. Никифоров, Н. Н. Макаров, Д. А. Ляхманов**

## МЕТОД ОБНАРУЖЕНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ С ПОЗИЦИЙ ТЕОРИИ АКТИВНОГО ВОСПРИЯТИЯ

Нижегородский государственный технический университет им. Р. Е. Алексеева

Задача обнаружения объектов на изображении или в видеопотоке – одна из актуальных задач компьютерного зрения. Ключевой этап в решении данной проблемы – формирование признакового описания объекта. Предлагается новый подход к формированию признакового описания объекта на основе теории активного восприятия. Приводятся результаты тестирования предложенного метода на базе данных изображений.

*Ключевые слова:* обнаружение объектов, теория активного восприятия, распознавание образов.

### Введение

Задача обнаружения объектов на изображении – одна из фундаментальных проблем в области компьютерного зрения и обработки изображений. Практическое применение данная задача находит в системах автопилотирования, индексирования изображений и видео, охранных системах.

Структуру системы обнаружения объектов на изображении можно представить в виде совокупности трёх блоков: предварительная обработка изображения, формирование признакового описания, принятие решения.

Предварительная обработка изображения обычно заключается в применении к изображению фильтра, подавляющего шум. Часто реализация данного этапа не выполняется, в этом случае ответственность за устойчивость к шуму перекладывается на метод формирования признакового описания.

При формировании признакового описания объекта используется гистограмма ориентированных градиентов, каскады Хаара и т. д.

Блок принятия решения заключается в формировании модели на основе признакового описания с использованием нейронных сетей, метода опорных векторов,  $k$ -ближайших соседей.

В [1] рассмотрена задача детектирования объекта, представляющего собой несколько соединенных сфер одинакового радиуса. В решении поставленной задачи определен ряд этапов: выделение границ объекта с использованием детектора Канни, поиск структурных элементов на основе преобразования Хафа, определение положения объекта с использованием распознавания по эталону и поиска изменений по кадрам. К недостаткам предложенного метода распознавания по эталону можно отнести частный характер задачи (должна быть известна структура объекта), значительные затраты времени и вычислительных ресурсов на формирование базы данных всевозможных изображений объекта.

В работе также рассмотрен подход к решению задачи определения положения объекта с использованием нейросети. К достоинствам использования неокөгнитрона при распознава-

нии относятся отсутствие необходимости этапов поиска границ и выделения окружностей, а также возможность внесения изменений для каждой новой задачи. К недостаткам следует отнести сложность структуры сети, порождающую большой объем вычислений и отсутствие возможности самообучения.

В [2] предложен метод распознавания, относящийся к классу обобщающих методов. Объект рассматривается как совокупность элементов, в качестве которых выбраны линии границ. Для обучения используется эталонное изображение объекта, на основе которого строится эталонный каркас. Исследуемое изображение обрабатывается фильтром Хаара с последующим выделением и соединением точек. Обнаружение заключается в построении каркаса исследуемого изображения, поиск наилучшего совмещения с эталоном, принятие решения об обнаружении путем сравнения количества совпавших линий с пороговым значением. К недостаткам данного метода можно отнести то, что он применим только для ограниченного класса объектов характерной формы.

В [3] описан подход к распознаванию объекта на изображении с использованием алгоритма адаптивного усиления (AdaBoost), в основе которого лежит идея отбора и комбинирования слабых классификаторов. Тестирование реализации предложенного алгоритма показало приемлемое время обработки изображений в тренировочной выборке. Установлено также, что в реальных условиях требуется обучающая выборка большого объема, а на обучение каскада классификаторов необходимо потратить несколько дней.

В [4] описан метод детектирования лиц на изображениях, в основу которого положены идеи алгоритма Viola&Jones. Для повышения скорости и точности детектирования в предобработке изображения предложено использовать серию фильтров, где каждый следующий фильтр обрабатывает только «перспективные» части изображения, полученные от предыдущих фильтров. При этом на основе использования информации о цвете и форме объекта, достигается отделение регионов, не содержащих лиц. Для детектирования лиц использовались совокупности анизотропных гауссовых примитивов, объединенные в каскадную модель (AdaBoost). Приведённые результаты экспериментов показывают, что предложенный гибридный метод превосходит классические по качеству и скорости распознавания, однако, сохраняются недостатки, связанные с трудоемкостью обучения каскада классификаторов.

В [5] представлен алгоритм обнаружения объектов, основанный на обучении свёрточной нейронной сети. Особенностью алгоритма является то, что в единой сети решаются сразу две задачи: выделение прямоугольных блоков, содержащих объекты, и определение принадлежности объекта некоторому классу. Для распознавания исходное изображение приводится к размеру  $448 \times 448$ , делится на ячейки  $7 \times 7$ , для каждой ячейки формируется метка, характеризующая ее принадлежность объекту некоторого класса.

Достоинством предложенного алгоритма является высокая скорость обработки изображений. К недостаткам следует отнести затраты на обучение сети, снижение точности обнаружения по сравнению с аналогичными алгоритмами, ограничения, связанные с размерами ячеек сетки (возникают трудности при обнаружении мелких объектов и объектов, расположенных близко друг к другу).

Данная работа посвящена решению задачи обнаружения объектов на изображении с позиций теории активного восприятия (ТАВ). Использование ТАВ позволит распознавать объекты произвольных классов с приемлемой трудоемкостью и высокой точностью, за счёт нового подхода к решению задачи формирования признаков.

### **1. Формирование признакового описания объекта с позиций теории активного восприятия**

Базовым преобразованием ТАВ является  $U$ -преобразование, которое реализуется в два этапа [6]. На первом этапе к изображению применяется  $Q$ -преобразование, после которого получаем матрицу визуальных масс  $m$  размером  $4 \times 4$  элемента:

$$m[i, j] = \sum_{k=\left(\frac{i-1}{4}\right) \cdot N}^{\frac{i}{4} \cdot N} \sum_{l=\left(\frac{j-1}{4}\right) \cdot M}^{\frac{j}{4} \cdot M} I[k, l], i = \overline{1, 4}, j = \overline{1, 4},$$

где  $I$  – изображение размером  $N \times M$  отсчётов.

Для корректной интерпретации результатов  $Q$ -преобразования отсчёты изображения должны принадлежать положительной области значений. С позиций ТАВ $Q$ -преобразование соответствует этапу предварительной обработки изображения (рис. 1).

На втором этапе к результату  $Q$ -преобразования применяется множество фильтров  $F = \{F_i\}$ ,  $i = \overline{1, 16}$  (рис. 1). В результате, формируется вектор спектральных коэффициентов  $\mu$ :

$$\mu_i = \sum_{k=1}^4 \sum_{l=1}^4 m[k, l] \cdot F_i[k, l].$$

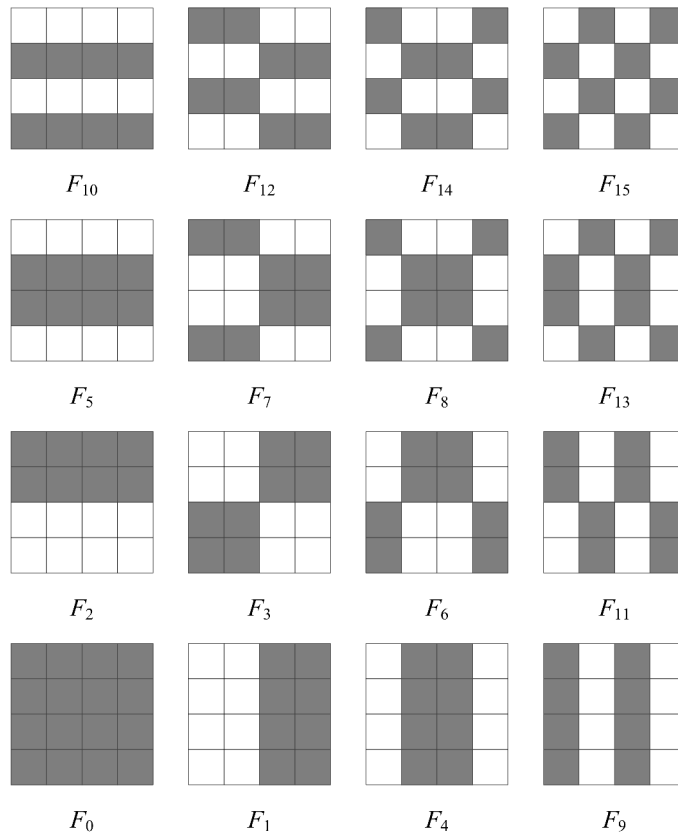


Рис. 1. Фильтры, используемые при вычислении признакового описания

Размер каждого фильтра составляет  $4 \times 4$  элемента. Элемент фильтра может принимать значения «+1» (тёмные области на рис. 1) и «-1» (светлые области на рис. 1). Конструктивно данные фильтры подобны фильтрам Уолша системы Хармута. Специфика использования данных фильтров заключается в том, что они применяются после реализации  $Q$ -преобразования.

## 2. Метод обнаружения объектов

Информационная модель обнаружения объекта на изображении показана на рис. 2.

Этап обучения заключается в выполнении следующих шагов:

1) формируются «положительные» и «отрицательные» образцы объектов для каждого из обнаруживаемых классов объектов; база данных «положительных» образцов включает изображения обнаруживаемых  $C$  классов объектов, каждый образец имеет размер  $h_c \times w_c$ ,

где  $h_c$  – высота образца;  $w_c$  – ширина образца;  $c$  – класс образца ( $c \in \overline{1, C}$ ), причём каждый образец содержит изображение детектируемого объекта, база данных «отрицательных» образцов строится аналогично базе для «положительных», каждый «отрицательный» образец не содержит изображение детектируемого объекта;

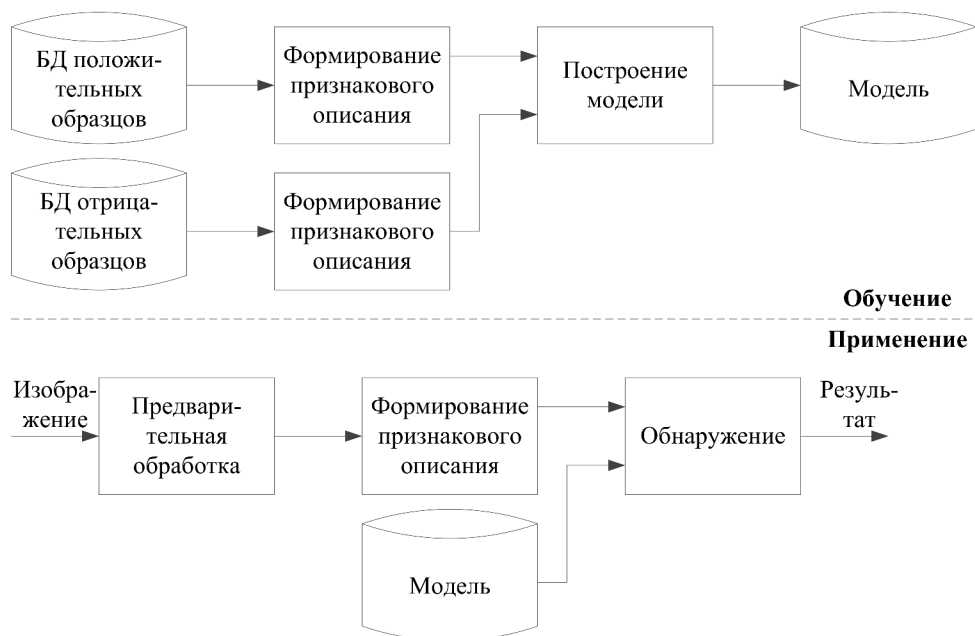


Рис. 2. Информационная модель обнаружения объектов

2) вычисляется признаковое описание для изображений, находящихся в указанных базах данных; формирование признакового описания изображения  $I$  выполняется по следующему алгоритму:

2.1) изображение  $I$  разбивается на  $L^2$  равных частей без перекрытия  $\mathbf{P}=\{P_i\}$ ,  $i=\overline{1, L^2}$  размером  $(N/L) \times (M/L)$  отсчётов (эквивалентность частей изображения выбрана исходя из того, что изображение обрабатывается в условиях априорной неопределённости), где  $P_i$  –  $i$ -я область изображения;  $L$  – количество разбиений одной стороны изображения;

2.2) для каждой части вычисляется признаковое описание на основе  $U$ -преобразования:

$$D = FC [I],$$

где оператор  $FC[\bullet]$  вычисляет признаковое описание изображения и реализуется следующим образом:

$$\forall j = \overline{1, L^2} : \\ D = D \cup U[P_j],$$

где  $U[\bullet]$  – оператор вычисления  $U$ -преобразования;  $D$  – признаковое описание изображения  $I$ .

Таким образом, число частей, на которое разбивается изображение, определяет размерность признакового описания:  $1 \times (15 \cdot L^2)$ . В результате выполнения данного шага формируется признаковое описание положительных образцов для каждого класса  $F_P = \{F_{P,i}\}$  и признаковое описание отрицательных образцов для каждого класса:  $F_N = \{F_{N,i}\}$ ,  $i = \overline{1, C}$ ;

3) с использованием метода опорных векторов на основе вычисленных признаковых описаний формируется модель для детектирования каждого из классов объектов; таким образом, каждый классификатор обучается различать два класса: «положительный» образец некоторого класса и «отрицательный» образец для того же класса:

$$\mathbf{S} = \{S_i\}, S_i = SVM_i [F_{P,i}, F_{N,i}], i = \overline{1, C},$$

где  $SVM[\bullet]$  – оператор определения параметров модели метода опорных векторов;  $S$  – множество моделей, полученных в результате обучения классификатора для каждого из классов объектов;

Этап применения заключается в выполнении следующих шагов:

1) предварительная обработка изображения  $I$  – деление изображения  $I$ , на котором выполняется детектирование объектов, на области с шагом  $s_h$  по горизонтали и  $s_v$  по вертикали, размер области для  $i$ -го класса объектов известен заранее и установлен на этапе обучения, выделение областей выполняется на нескольких масштабах; это используется для обнаружения объектов, размер которых меньше размера эталона:

$$Q = \{Q_{i, (x, y), s}\}, i = \overline{1, T},$$

где  $Q$  – множество всех областей, полученных по изображению;  $Q_{i, (x, y), s}$  –  $i$ -я область  $(x, y)$  – координата верхнего правого угла области;  $s$  – масштаб изображения ( $0 < s < 1$ );  $T$  – количество полученных областей;

2) для каждой области  $Q_i$  формируется признаковое описание:

$$D_{Q_{ixys}} = FC [Q_{i, (x, y), s}], i = \overline{1, T},$$

где  $D_{Q_{ixys}}$  – признаковое описание области  $Q_{i, (x, y), s}$ ;

3) при выполнении обнаружения на вход классификатора направляются полученные на предыдущем этапе признаковые описания областей изображения:

$$C_{Q_{ixys}} = SVM_u [D_{Q_{ixys}}, S],$$

где  $SVM_u$  – оператор определения класса по признаковому описанию  $D_{Q_{ixys}}$  на основе множества моделей  $S$ ;  $C_{Q_{ixys}}$  – класс области  $Q_{i, (x, y), s}$ ;

4) учитывая, что на предыдущем шаге для одного объекта, находящегося на тестовом изображении, генерируется множество близких друг другу вариантов расположения («гипотез»), необходимо выполнить подавление немаксимумов с использованием алгоритма Soft-NMS, описанного в [7]; данный метод выбирает область с максимальной оценкой правдоподобия некоторого класса объектов, а все остальные области, перекрывающиеся в определённом процентном соотношении с данной областью, подавляет.

### 3. Вычислительный эксперимент

Вычислительный эксперимент предназначен для оценки точности работы предложенного метода детектирования объектов. Для проведения вычислительного эксперимента были выбраны объекты двух классов – «Пешеход» и «Легковой автомобиль». В обучающей выборке для каждого из классов хранится 1000 положительных и 1000 отрицательных эталонов, таким образом, всего 4000 изображений. Тестовая выборка включает 1000 предварительно размеченных изображений для каждого из классов размером 640×480 отсчётов.

Размер эталона для класса «Пешеход» составляет 64×128 отсчётов, для класса «Легковой автомобиль» – 64×64 отсчёта. Разметка изображений выполнялась в программе VGGImageAnnotator, разработанная группой визуальной геометрии Оксфордского университета (см. рис. 3). На этапе применения величина смещения окна по изображению кратна значению  $L$ .

В табл. 1 приводятся результаты тестирования предложенного метода обнаружения объектов.

Таблица 1

Результаты вычислительных экспериментов

$L$	Класс «Пешеход»		Класс «Легковой автомобиль»	
	Ошибка первого рода (%)	Ошибка второго рода (%)	Ошибка первого рода (%)	Ошибка второго рода (%)
1	10	15	9	11
2	7	5	6	4
4	2	3	3	3

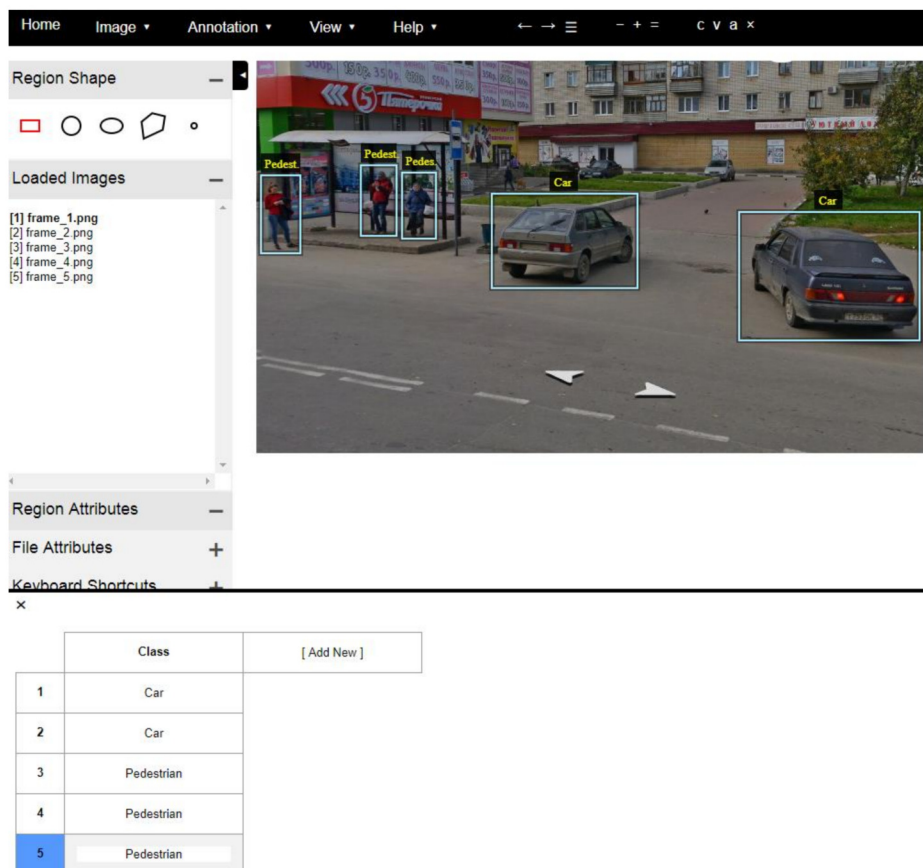


Рис. 3. Разметка тестового изображения

На рис. 4 показан результат детектирования объектов класса «Пешеход» и «Легковой автомобиль». Время обработки одного изображения составляет 47 мс.



Рис. 4. Результат детектирования объектов

В работе [2] решается задача обнаружения жеста открытой ладони в видеопотоке. Обучающая выборка составляла одно полутоновое изображение  $640 \times 480$  отсчётов. Тестовая выборка – 52 изображения, содержащих ладонь, и 48 изображений, не содержащих ладонь. Точность обнаружения 89%, среднее время обработки одного изображения – 34 мс.

В работе [3] решается задача обнаружения бейсбольного мяча. В качестве обучающей выборки используется одно изображение  $150 \times 150$  отсчётов. Обучающая выборка составляет 1000 изображений (500 – содержат объект, 500 – не содержат объект). Время обработки одного изобраа-

жения – не более 13 мс. Время, затраченное на обучение каскада классификаторов – 1 день.

В работе [4] решается задача обнаружения лиц на изображении. Тестирование производилось на специализированном тестовом наборе CMU/MITTestset, состоящем из 130 изображений, содержащих в общей сложности 507 лиц. Ошибка детектирования составляет 6,1%, время на обработку одного изображения – 35 мс.

Таким образом, реализация предложенного метода обнаружения объектов по точности обнаружения и производительности не уступает реализациям известных методов;

### Заключение

Методы обнаружения объектов на изображении имеют широкий круг применения, включая робототехнику, анализ медицинских изображений, охранные системы. В данной работе предложен метод обнаружения объектов, разработанный с позиций теории активного восприятия. Выполнено тестирование предложенного метода на двух классах объектов – «Пешеход» и «Легковой автомобиль». Результаты тестирования подтвердили эффективность предложенного метода обнаружения. По точности обнаружения объектов, предложенный метод не уступает известным подходам, а в ряде случаев – показывает лучшие результаты. Цель дальнейших исследований – повышение устойчивости предложенного метода обнаружения к искажениям анализируемого изображения.

### Библиографический список

1. Азаренко, Д.С. Детектирование объекта на изображении и определение его смещения на двух различных изображениях // Искусственный интеллект. – 2013. – №3. – С. 90–97.
2. Алфимцев, А.Н. Метод обнаружения объекта в видеопотоке в реальном времени / А.Н. Алфимцев, И.И. Лычков // Вестник ТГТУ. – 2011. – Т. 17. – № 1. – С. 44–55.
3. Бутенко, В. В. Поиск объектов на изображении с использованием алгоритма адаптивного усиления // Молодой ученый. – 2015. – №4. – С. 52–56.
4. Гребнов, И. В. Новый метод детектирования человеческих лиц на цифровых изображениях // Вестник ИГЭУ. – 2008. – Вып. 4. – С. 77–81.
5. Redmon, J. The IEEE / J. Redmon [et al.] // Conference on Computer Vision and Pattern Recognition (CVPR), 2016. – P. 779–788.
6. Utrobin, V. A. Physical interpretation of the elements of image algebra // Phys. Usp. 47 1017–1032 (2004).
7. Bodla, N. Improving Object Detection With One Line of Code / N. Bodla [et al.] // arXiv preprint arXiv:1704.04503. – 2017.

*Дата поступления  
в редакцию 31.01.2018*

**V.E. Gai, V.A. Utrobin, E.N. Vikulova, N.A. Nikiforov, N. N. Makarov, D.A. Lyakhmanov**

### THE OBJECT DETECTION METHOD IN AN IMAGE IN TERMS OF THE THEORY OF ACTIVE PERCEPTION

Nizhny Novgorod state technical university n.a. R.E. Alekseev

**Purpose:** A new approach to the formation the object parametric description in order to build object detection systems in an image or a video stream.

**Design/methodology/approach:** The parametric description forming includes two stages: a pre-processing stage, a filtering stage and obtaining a spectral coefficient vector. The basic transformations of the theory of active perception is the basis for both stages. A training phase is conducted for the purpose of object detection, including the formation of parametric descriptions for positive and negative samples of each particular object class. A detecting phase includes image preprocessing, object feature description forming and classification.

**Experiments:** The proposed method was evaluated for detecting objects belonging to two classes. The test results proved the effectiveness of the proposed detection method. The proposed method doesn't concede to known approaches in terms of the object detection accuracy and demonstrates better results in some cases.

**Findings:** The proposed method of object detection in the image has a wide range of applications, including robotics, medical image analysis, security systems.

**Research implications:** The purpose of further research – improving the proposed detection method stability of distorted images.