

УДК 004.932.72

DOI:10.38028/ESI.2022.28.4.016

## Система распознавания объектов с использованием теории активного восприятия

Гай Василий Евгеньевич, Смирнов Александр Вадимович, Баринов Роман Олегович, Поляков Игорь Владимирович, Голубенко Владимир Андреевич, Кузнецов Георгий Дмитриевич

Нижегородский государственный технический университет им. Р.Е. Алексеева, Россия, Нижний Новгород, [barinovr@list.ru](mailto:barinovr@list.ru)

**Аннотация.** Предлагается метод обнаружения объекта на изображении на основе глобального признакового описания. Описывается информационная модель и предлагаются варианты реализации каждого из этапов решения задачи обнаружения объекта на изображении. На этапе предварительной обработки изображения приводятся варианты, включающие нормализацию, вычисление функции яркости и применение фильтра Гаусса. Этап формирования глобального признакового описания объекта опирается на теорию активного восприятия (U-преобразование). Принятие решения о локализации объекта осуществляется на основе минимального евклидова расстояния от обнаруженного объекта до эталонных объектов из базы данных. Для тестирования использовались изображения из базы Russian Traffic Sign Dataset и их изменённые копии (изображения с наложенным шумом, изображения с поворотом искомым объектов). При анализе результатов тестирования были предложены параметры, дающие наибольшую точность предложенного метода обнаружения объекта на изображении. При наличии шума на изображении, точность локализации предложенного метода составила более 70%. Предлагаемый метод обнаружения объекта на изображении устойчив к повороту искомым объектов. Полученная точность, порядка 94-96%, при сравнении с точностью существующих методов показала, что в нормальных условиях разработанный метод работает аналогично известным методам.

**Ключевые слова:** обработка изображений, обнаружение объектов, теория активного восприятия

**Цитирование:** Гай В.Е. Система распознавания объектов с использованием теории активного восприятия / В.Е. Гай, А.В. Смирнов, Р.О. Баринов, И.В. Поляков, В.А. Голубенко, Г.Д. Кузнецов // Информационные и математические технологии в науке и управлении. – 2022. – № 4(28). – С. 201-212. – DOI:10.38028/ESI.2022.28.4.016.

**Введение.** Распознавание объектов на изображениях стало в последнее время одной из важнейших задач анализа сцен и машинного зрения. Под объектом понимается не только цифровое представление локального фрагмента двумерной сцены, а некоторое его приближенное описание, в виде набора характерных свойств (признаков). Основное назначение описаний (образов объектов) – это их использование в процессе установления соответствия объектов, осуществляемого путем сравнения (сопоставления).

Исследования по распознаванию образов пространственных объектов отличаются большим разнообразием в постановке задач и выборе средств их решения (методов обработки соответствующих фрагментов изображений), что является следствием разнообразия областей практического применения. Традиционными задачами, решавшимися еще в первых опытных разработках систем машинного зрения, служат задачи обнаружения и распознавания объектов, имеющих заданную форму на основе искаженных (зашумленных, повернутых, деформированных) изображений.

**1. Структура алгоритмов обнаружения объектов на изображении.** Большая часть актуальных методов, решающих задачу обнаружения объектов на изображении, включают в себя три основных этапа:

- предварительная обработка изображения;
- построение признакового описания искомого объекта;
- принятие решения на основе полученного признакового описания.

Схема описанных выше этапов решения задачи представлена на рисунке 1.



Рис. 1. Структура системы обнаружения объектов на изображении

**2. Обзор существующих методов построения признакового описания объекта на изображении.** Общие принципы работы существующих систем по обнаружению объектов достаточно схожи. Как правило, в процессе поиска производится сравнение признаков данного объекта или группы данных объектов в используемой базе изображений. Степени схожести объектов задаются параметрами системы и задачами, которые она выполняет. Главное преимущество глобальных дескрипторов заключается в том, что такого рода признаки демонстрируют достаточно высокую производительность, при этом обладая существенно меньшей сложностью вычислений.

Для получения глобального признакового описания, используются низкоуровневые операции, которые производятся при структурном анализе изображения. Благодаря этому, такого рода дескрипторы могут нести информацию о текстуре объекта, его цвете, или возможном движении объекта.

Сегодня имеются несколько алгоритмов, которые применяются для вычисления глобального признакового описания объектов. Ниже приведены самые основные и наиболее популярные из них: GIST [1]; HOG [2]; SOHOG [3]; Моментные инварианты Ху [4].

GIST – алгоритм базируется на допущении, что при помощи распределения градиентов интенсивности или направления краев можно описать внешний вид и форму объектов на изображении. Реализация такого рода дескриптора может осуществляться разделением изображения на сегментированные области, именуемые ячейками, и расчетом для каждой ячейки гистограммы направлений градиентов или направлений краев для пикселей, находящихся внутри ячейки. Главным недостатком такого метода является сложность вычислений направлений градиентов и направлений краев пикселей.

HOG – этот метод похож на гистограммы направления края, дескрипторы SIFT и контексты формы, но отличается тем, что вычисляется на плотной сетке равномерно распределенных ячеек по всему изображению и использует нормализацию перекрывающегося локального контраста для увеличения точности. Недостатком метода HOG является применение операции свертки при вычислении градиентов направлений.

SOHOG – совместная гистограмма ориентированных градиентов, представляет собой дескриптор, основанный на наборе из нескольких пар допустимых градиентных ориентаций. SOHOG может выражать формы более подробно, чем HOG, который использует ориентацию с одним градиентом. Недостатком данного метода является повышенная вычислительная сложность по отношению к стандартному методу HOG.

Моментные инварианты Ху – хорошо известный подход в распознавании образов на цифровом изображении. Приведенный Ху набор характеристик включает семь инвариантов. Алгоритм [5] получения инвариантов содержит следующие этапы:

1. Определение центральных моментов порядка не выше третьего.
2. Получение моментов, инвариантных к операциям поворота, переноса и зеркального отображения [6].

### 3. Получение моментов, инвариантных к полной группе аффинных преобразований.

Как показано в работе Ху, могут быть построены инварианты относительно перемещения, масштаба и вращения. Следует заметить, что описанные Ху инварианты являются такими только в непрерывной области. В дискретной области ни масштабирование, ни вращение не определены четко: дискретное изображение, преобразованное таким образом, обычно является приближенным – это является основным недостатком описанного метода.

**3. Информационная модель обнаружения объектов на изображении.** Предлагаемый метод решения задачи обнаружения объекта на изображении комбинирует метод локализации объекта по цвету в цветовом пространстве HSV на этапе детектирования объекта с применением теории активного восприятия (ТАВ) [7, 8] на этапе формирования глобального признакового описания выделенного объекта. На этапе принятия решения о распознанном на изображении объекте применяется алгоритм поиска наименьшего евклидова расстояния между глобальным признаковым описанием обнаруженного объекта и базой эталонных признаковых описаний известных объектов в пространстве признаков.

В качестве распознаваемых объектов в данной работе будут использоваться дорожные знаки. Таким образом, база эталонов будет включать в себя непосредственно изображения самих дорожных знаков, а база данных анализируемых изображений будет сформирована из изображений, покадрово полученных с видеорегистратора, находящегося в салоне автомобиля. Системе будет необходимо проанализировать входное изображение, выделить на нем область, в которой присутствует дорожный знак, а затем – распознать, какой именно знак представлен на изображении.

**3.1. Предварительная обработка изображения.** Цель предварительной обработки изображения – избавление от шумов и сглаживание цветов на изображении. Данная операция позволяет значительно повысить точность на этапе локализации.

В первую очередь, к изображению применяется фильтр Гаусса, он позволяет избавиться от большинства шумов. Хотя это и приводит к некоторому размытию изображения, цветовые диапазоны сглаживаются и выравниваются.

Фильтр Гаусса – это фильтр размытия изображения, который использует нормальное распределение для вычисления преобразования, применяемого к каждому пикселю изображения. Нормальное распределение для двух измерений описывается формулой [9]:

$$G(u, v) = \frac{1}{2\pi r^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

где  $r$  – радиус размытия,  $\sigma$  – стандартное отклонение распределения Гаусса.

Данная формула задает поверхность, имеющую вид концентрических окружностей с нормальным распределением от центральной точки. Пиксели, где распределение отлично от нуля, используются для построения матрицы свертки, которая применяется к исходному изображению. Значение каждого пикселя становится средне взвешенным для окрестности. Исходное значение пикселя принимает наибольший вес (имеет наивысшее Гауссово значение), и соседние пиксели принимают меньшие веса, в зависимости от расстояния до них.

Параметры для фильтра Гаусса были подобраны эмпирически и представляют собой ядро размером 5 на 5 пикселей, применяемое к изображению в три последовательных прохода.

**3.2. Детектирование объекта на изображении.** Специфические условия освещённости значительно влияют на возможность корректного восприятия цвета объектов снимаемой сцены. При съёмке реальной дорожной ситуации на знаках возникает целый ряд различных условий освещения, сильно усложняющих детектирование объектов: прямые солнечные лучи, отражённый свет, затенения, свет автомобильных фар в тёмное время суток. Таким образом, на изображениях, получаемых с камеры, установленной в автомобиле, не всегда воз-

можно выделить интересующую область по цвету с помощью простого применения порогового фильтра напрямую в цветовом пространстве RGB (Red, Green and Blue). Для извлечения красного, синего и желтого цвета из входного изображения необходимо использовать информацию о цвете каждого пикселя независимо от его неконтролируемого показателя освещенности. Для этих целей было выбрано цветовое пространство HSV (Hue, Saturation and Value). Цветовое пространство HSV принадлежит к группе цветовых пространств, изначально основанных на человеческом восприятии цветов.

Таким образом, следующим этапом после предварительной обработки является наложение маски, фильтрующей заданный цветовой диапазон в HSV пространстве (синий, красный и желтый – основные цвета дорожных знаков). Однако, в большинстве случаев, наложение маски приведет не только к выделению синих, красных и желтых знаков на изображении, но также и к выделению других регионов изображения, содержащих цвета заданного цветового диапазона. Таким образом, необходимо снизить погрешность модуля локализации путем устранения нежелательных регионов, выделенных цветовой фильтрацией по заданной маске, но так, чтобы случайно не отсеять интересующие регионы изображения, содержащие дорожные знаки.

В настоящей работе с этой целью применяются морфологические операции над выделенными регионами. Наиболее подходящими морфологическими операциями являются операции дилатации и эрозии. Операция дилатации служит для увеличения, «наращивания» контура вокруг выделенного маской региона, а операция эрозии – напротив, служит для уменьшения, «сужения» контура вокруг выделенного маской региона. Таким образом, для устранения не интересующих регионов, первой применяется операция эрозии в два последовательных прохода, а затем – дилатация в четыре последовательных прохода.

Также на этапе локализации к изображению применялся алгоритм Рамера-Дугласа-Пекера [10], позволяющий уменьшать число точек на кривой, которая аппроксимирована большей выборкой точек. Применение этого алгоритма позволяет фильтровать контуры по форме самих контуров. Таким образом, вначале задается ряд контуров, описывающих интересующие формы контуров, повторяющие формы дорожных знаков – круг, квадрат, треугольник. Затем, с помощью указанного алгоритма каждый из обнаруженных на изображении контуров аппроксимируется и проверяется на схожесть с одним из шаблонных контуров. Если наблюдается совпадение с одним из шаблонов, контур помечается как искомый и остается в выборке контуров, если совпадение ни с одним из шаблонов не наблюдается – контур удаляется из выборки контуров.

**3.3. Формирование глобального признакового описания изображения с помощью теории активного восприятия.** Исследуемое изображение представлено в виде функции яркости:

$$I(x, y), x \in X, y \in Y,$$

где  $X$  и  $Y$  – конечные множества. В результате применения к такому изображению  $U$ -преобразования вычисляется спектральное представление изображения:

$$D = \{d_i\}, i = \overline{1, M},$$

где  $M$  – число сегментов, на которые было разбито исследуемое изображение;  $d_i$  – спектральное представление сегмента  $i$ , включающее  $L$  спектральных коэффициентов ( $L$  – число фильтров, используемых на этапе дифференцирования);  $d_i\{k\}$  –  $k$ -й спектральный коэффициент представления сегмента  $i$ .

Формирование исходного описания с позиции теории активного восприятия изображения понимается как операция интегрирования, и включает в себя два следующих этапа:

- нормирование изображения;
- вычисление матрицы «визуальных масс».

Нормирование изображения происходит путем вычитания из каждого элемента матрицы минимального элемента матрицы, а затем деления каждого элемента матрицы на максимальный элемент матрицы.

Для получения матрицы «визуальных масс» необходимо выполнить так называемое в теории активного восприятия, Q-преобразование.

Q-преобразование выполняется по формуле:

$$B_{kl} = \sum_{i=1}^2 \sum_{j=1}^2 A_{kl}[i, j], k = \overline{1,4}, l = \overline{1,4}.$$

Формирование системы признаков с позиции теории активного восприятия изображения понимается как операция дифференцирования. Результатом операции дифференцирования является вектор «спектральных коэффициентов»  $\mu = (\mu_0, \mu_1, \mu_2, \dots, \mu_{15})$ . Операция дифференцирования реализуется по средствам 16 фильтров (рисунок 2), применяемых к матрице визуальных масс, где темный элемент означает умножение исходного элемента матрицы визуальных масс на -1, светлый элемент – умножение исходного элемента матрицы визуальных масс на +1. Все полученные в результате применения фильтра к матрице значения суммируются между собой, а результирующая сумма называется спектральным коэффициентом.

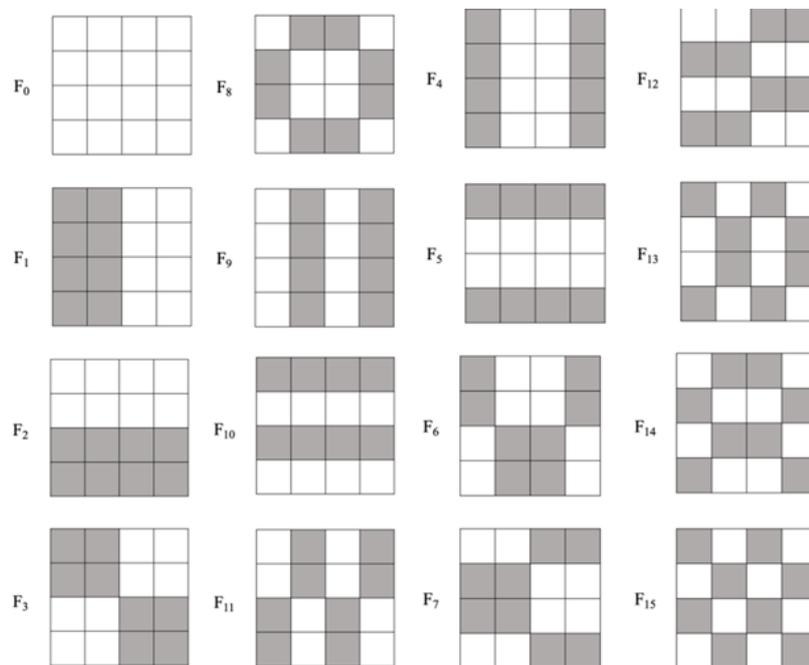


Рис. 2. Фильтры, использующиеся для вычисления U-преобразования

**3.4. Принятие решения.** Этап принятия решения заключается в получении соответствия между обнаруженным объектом и одним из эталонных объектов, известных системе. С этой целью в пространство признаков, полученное по глобальному признаковому описанию объектов-эталонов, вносится глобальное признаковое описание объекта, полученное после его обнаружения на изображении. Далее, необходимо вычислить ближайшее признаковое описание эталонного объекта к обнаруженному объекту. В качестве меры близости целесообразно использовать евклидово расстояние:

$$(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}.$$

В результате вычисления евклидова расстояния от глобального признакового описания обнаруженного объекта до глобального признакового описания каждого из эталонных объ-

ектов получается числовой вектор, минимальный элемент которого представляет собой искомое минимальное евклидово расстояние. Этот элемент указывает на эталонный объект, который оказался в признаковом пространстве ближе всего к обнаруженному объекту.

**4. Вычислительный эксперимент.** Для проведения эксперимента была сформирована база изображений объектов. Как было отмечено выше, в роли объектов выступают дорожные знаки. В базе содержатся изображения 156 различных дорожных знаков, в качестве которых взяты наиболее основные и часто встречающиеся в реальных дорожных условиях. Каждый объект в базе имеет четыре шаблона, которые используются непосредственно при обнаружении данного объекта на изображении. Каждый из шаблонов такого рода отличается погодными условиями и условиями освещения, при которых проведена съемка объекта.

Для проведения вычислительного эксперимента была использована Российская база изображений автодорожных знаков [11], известная так же, как Russian Traffic Sign Dataset (RTSD). Примеры кадров из этой базы представлены на рисунке 3. Данная база содержит:

- 179138 размеченных кадров с авторегистратора, содержащих дорожные знаки;
- 156 типов дорожных знаков;
- 104358 изображений знаков (полученных путем кадрирования).

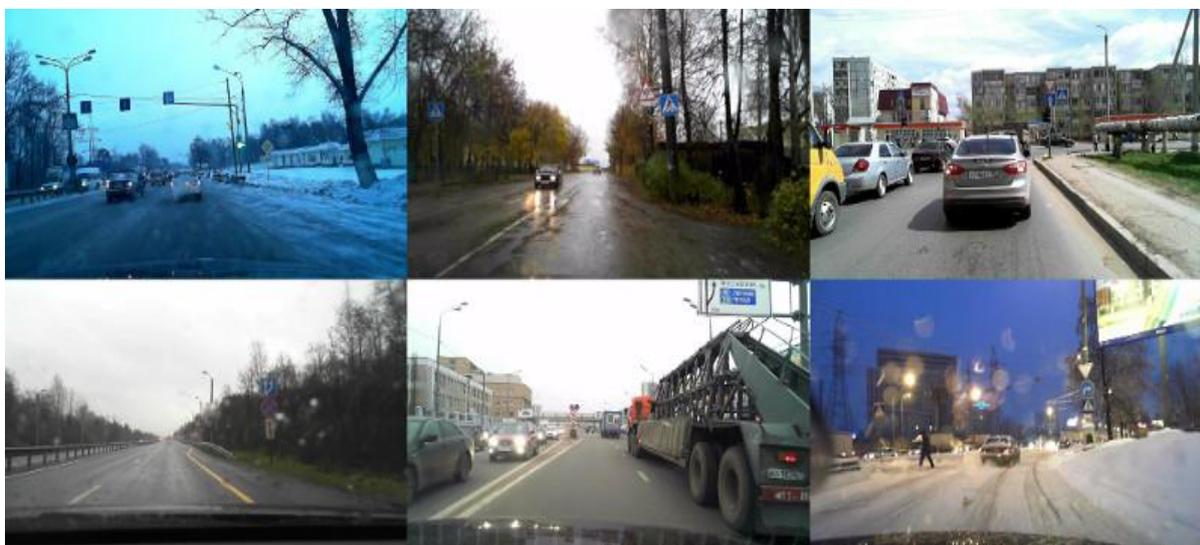


Рис. 3. Примеры кадров из базы изображений RTSD

**4.1. Постановка вычислительного эксперимента.** Для проведения вычислительного эксперимента из базы RTSD случайным образом были отобраны 10000 изображений, сделанных в разное время суток, при разной погоде, в разные сезоны года.

Кроме того, для усложнения задачи и внесения дополнительной объективности в проведение эксперимента и апробацию предлагаемого подхода, были искусственно созданы две дополнительные выборки изображений на основе базы RTSD. Первая синтезированная выборка состоит из изображений, к которым дополнительно добавлен шум различной степени. Вторая выборка включает в себя изображения, дорожные знаки на которых повернуты под различными углами. Каждая из выборок включает в себя 10000 изображений. Таким образом, итоговые выборки для проведения вычислительного эксперимента сформированы следующим образом:

- Выборка изображений из базы RTSD;
- Выборка изображений из базы RTSD с добавлением шума различной интенсивности: ОСШ = 20дБ, 10дБ, 0дБ;
- Выборка изображений из базы RTSD с поворотом искомых объектов: Угол поворота = 15°, 30°, 45°.

Оценка точности работы системы проводилась следующим образом: в первую очередь, оценивалась точность результатов работы локализатора с различными параметрами, без дальнейшей классификации. Цель – выявление оптимального набора параметров локализатора, при котором точность локализации максимальна. Далее, проводилась оценка точности классификации по следующему принципу: к локализатору применялся оптимальный набор параметров, оценка классификации выполнялась только при условии, что на этапе локализации не было допущено ошибок.

**4.2. Анализ результатов вычислительного эксперимента.** В таблицах, содержащих результаты работы локализатора, отражены варьируемые параметры, такие, как цветовой диапазон различных компонентов в пространстве HSV и количество итераций применения морфологических операций Erosion и Dilation к изображению, а также показатели точности локализации в процентах.

В таблице 1 приведены три варианта установки цветového диапазона для каждого из целевых цветов в цветовом пространстве HSV.

**Таблица 1.** Варианты установки границ цветového диапазона

Номер цветového диапазона	Синий цвет	Желтый цвет	Красный цвет
1 H	95-130	19-37	160-20
1 S	140-255	130-255	135-255
1 V	100-255	110-255	115-255
2 H	99-125	23-34	170-10
2 S	155-255	141-255	141-255
2 V	115-255	124-255	124-255
3 H	105-115	25-31	175-5
3 S	165-255	150-255	152-255
3 V	125-255	136-255	135-255

Для каждого из этих наборов проводился вычислительный эксперимент, в котором варьировалось количество итераций морфологических операций Erosion и Dilation.

В таблице 2 приведены результаты вычислительного эксперимента с локализатором, проведенного с изображениями из базы RTSD. В таблицу включены комбинации параметров, при которых итоговая точность локализации составляла не менее 70%.

**Таблица 2.** Результаты вычислительного эксперимента с локализатором для выборки из базы RTSD

Номер цветového диапазона	Итераций Erosion	Итераций Dilation	Точность
1	2	2	76.43%
1	3	4	79.64%
1	4	1	74.12%
2	3	4	89.78%
2	4	3	90.33%
2	4	2	94.71%
2	2	2	86.52%
3	2	3	83.46%
3	4	4	79.12%
3	3	4	81.83%

Из таблицы 2 видно, что наилучшей комбинацией параметров локализатора является второй (средний) цветовой диапазон из трех описанных выше, и следующие параметры ал-

горитмов применения морфологических операций – 4 итерации для Erosion, 2 итерации для Dilation.

Результаты вычислительного эксперимента с локализатором, проведенного с изображениями из базы RTSD с добавлением шума различной интенсивности, приведены в таблице 3. В таблицу включены комбинации параметров, при которых итоговая точность локализации составляла не менее 70%.

Как видно из таблицы 3, добавление искусственного шума к изображению ухудшает качество работы локализатора. С ростом значения отношения сигнала к шуму, точность локализации снижается. Однако, наилучший результат показывает второй (средний) цветовой диапазон с пятью итерациями Erode и тремя итерациями Dilate.

Результаты вычислительного эксперимента с локализатором, проведенного с изображениями из базы RTSD с поворотом искомым объектов, приведены в таблице 4. В таблицу включены комбинации параметров, при которых итоговая точность локализации составляла не менее 70%.

**Таблица 3.** Результаты вычислительного эксперимента с локализатором для выборки из базы RTSD с добавлением шума различной интенсивности

Номер цветового диапазона	Итераций Erosion	Итераций Dilation	Точность		
			ОСШ = 20дБ	ОСШ = 10дБ	ОСШ = 0дБ
1	4	2	90.11	85.21%	70.41%
2	2	2	89.81	86.52%	71.90%
2	5	3	91.26%	87.46%	72.91%
3	4	4	88.28%	81.12%	72.14%
3	3	4	87.92%	84.61%	70.76%

Как видно из таблицы 4, поворот изображения искомого объекта на любой из заданных углов никак не влияет на качество работы локализатора. Это обусловлено тем, что алгоритм локализации ориентируется только на цвет, а не на положение объекта в кадре. Стоит отметить, что результаты локализации с приведенными комбинациями параметров из таблицы 4 хорошо согласуются с результатами локализации из таблицы 2 – аналогичные наборы параметров выдают аналогичные результаты. Наилучший результат показывает второй (средний) цветовой диапазон с четырьмя итерациями Erode и двумя итерациями Dilate. Именно такой набор параметров локализатора был принят как оптимальный и использован в дальнейших вычислительных экспериментах с классификатором.

**Таблица 4.** Результаты вычислительного эксперимента с локализатором для выборки из базы RTSD с поворотом искомым объектов

Номер цветового диапазона	Итераций Erosion	Итераций Dilation	Точность		
			УП = 15°	УП = 30°	УП = 45°
2	4	3	90.33%	90.33%	90.33%
2	3	4	89.78%	89.78%	89.78%
2	4	2	94.71%	94.71%	94.71%
1	3	4	79.64%	79.64%	79.64%
3	2	3	83.46%	83.46%	83.46%

В таблицах, содержащих результаты работы классификатора, отражены показатели точности в процентах, уровень накладываемого шума в децибелах, угол поворота изображения объекта в градусах, а также отражено наличие либо отсутствие экспериментальной функциональности во время проведения эксперимента – АРДП (алгоритм Рамера-Дугласа-Пекера).

Результаты вычислительного эксперимента с классификатором, проведенного с изображениями из базы RTSD, приведены в таблице 5.

**Таблица 5.** Результаты вычислительного эксперимента с классификатором для выборки из базы RTSD

АРДП	Точность
Да	98.53%
Нет	96.07%

Как видно из таблицы 5, применение алгоритма Рамера-Дугласа-Пекера увеличивает точность результатов.

Результаты вычислительного эксперимента с классификатором, проведенного с изображениями из базы RTSD с добавлением шума различной интенсивности, приведены в таблице 6. Как видно из этой таблицы, с увеличением интенсивности накладываемого шума, снижается точность распознавания. При этом корреляция между результатами работы системы с включенным и выключенным АРДП остается такой же, как и в эксперименте, проводимом на изображениях без шума.

**Таблица 6.** Результаты вычислительного эксперимента с классификатором для выборки из базы RTSD с добавлением шума различной интенсивности

АРДП	ОСШ	Точность
Да	20 дБ	91.67%
Да	10 дБ	87.32%
Да	0 дБ	81.12%
Нет	20 дБ	89.71%
Нет	10 дБ	84.59%
Нет	0 дБ	78.97%

Результаты вычислительного эксперимента с классификатором, проведенного с изображениями из базы RTSD с поворотом искомых объектов, приведены в таблице 7.

**Таблица 7.** Результаты вычислительного эксперимента с классификатором для выборки из базы RTSD с поворотом искомых объектов

АРДП	УП	Точность
Да	15°	97.87%
Да	30°	97.34%
Да	45°	96.91%
Нет	15°	95.17%
Нет	30°	94.76%
Нет	45°	94.23%

Как видно из таблицы 7, наличие угла поворота вплоть до 45° не оказывает существенного влияния на точность распознавания. Это говорит о том, что глобальное признаковое описание на основе теории активного восприятия является инвариантным и к масштабу, и к повороту объекта.

Пример работы локализатора и классификатора представлен на рисунке 4. Розовым прямоугольником отмечена область, обнаруженная локализатором, результат работы классификатора представлен в дополнительном окне.



**Рис. 4.** Результат работы локализатора и классификатора

В таблице 8 представлены результаты работы известных методов локализации объектов на изображении. Сравнивая данные из таблицы 8 и полученные результаты тестирования алгоритма в различных условиях, можно сделать вывод о том, что разработанный метод имеет точность обнаружения объектов на изображении, сопоставимую с точностью рассмотренных известных методов, а в ряде случаев точность предложенного подхода оказывается выше. Также стоит отметить ряд преимуществ разработанного метода.

Во-первых, U-преобразование имеет минимально возможную вычислительную сложность, в отличие от стандартных преобразований, требующих реализации свертки, а на уровне весовых коэффициентов – операции арифметического умножения.

Во-вторых, описанные в пункте 2.3 преобразования – это операции с целыми числами, и поэтому исчезает проблема накопления ошибки при округлениях.

В-третьих, за счёт применения U-преобразования к изображению дорожного знака целиком, учитываются только низкочастотные компоненты изображения. В этом случае детали мелки, а шумы и помехи исчезают.

**Таблица 8.** Точность работы известных методов обнаружения объектов

Метод	Точность
GIST	93-96%
HOG	89-94%
CoHOG	95-98%
Моменты Ху	94-98%

**Заключение.** Был разработан алгоритм решения задачи поиска объекта на изображении, который заключается в новой комбинации подходов для решения задачи на разных ее этапах, а также использовании нового способа формирования глобального признакового описания на этапе формирования признакового описания. В данной работе таким подходом является использование инструментария теории активного восприятия.

При анализе результатов тестирования были выбраны параметры, дающие наибольшую точность работы системы. Полученная точность, порядка 94-96%, при сравнении с точностью существующих методов показала, что разработанный метод работает аналогично известным методам. Кроме того, включая дополнительную функциональность в виде алгоритма Рамера-Дугласа-Пекера, предоставляется возможным повысить точность распознавания на 3-5%.

#### Список источников

1. Oliva A., Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vis., 2001, vol. 42, no. 3, pp. 145-175.

2. Dalal N., Triggs B. Histograms of oriented gradients for human detection. In CVPR, 2005.
3. Watanabe T., Ito S., Yokoi K. "Co-occurrence histograms of oriented gradients for pedestrian detection". PSIVT, 2009.
4. Hu M.K. Visual Pattern recognition by moment invariants, IRE trans. Info. Theory, 1962, vol. IT-8, pp.179-187.
5. Абрамов Н.С. Определение расстояний на основе системы технического зрения и метода инвариантных моментов / Н.С. Абрамов, В.П. Фраленко // Информационные технологии и вычислительные системы, 2012. – № 4. – С. 32-39.
6. Роджерс Д. Математические основы машинной графики / Д. Роджерс, Дж. Адамс // М.: Машиностроение, 1980. – 240 с.
7. Утробин В. А. Элементы теории активного восприятия изображений / В. А. Утробин // НГТУ, 2001. – 64 с.
8. Kondratiev V.V., Zhevnerchuk D.V. Application of methods of self-organization theory to problems of profiling and configuring computational systems. Doklady Mathematics, Pleiades Publishing, Ltd., 2014, vol. 90, no. 3, pp. 788-190, ISSN 1064-5624.
9. Haddad R.A., Akansu A.N. A class of fast gaussian binomial filters for speech and image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1991, vol. 39, pp 723-727.
10. Ramer Urs. An iterative procedure for the polygonal approximation of plane curves. Computer graphics and image processing, 1972, no. 1(3), pp. 244-256, DOI: 10.1016/S0146-664X(72)80017-0.
11. Шахуро В. Российская база изображений автодорожных знаков / В. Шахуро, А. Конушин // Компьютерная оптика, 2016.

*Гай Василий Евгеньевич. Инженер, доцент НГТУ им. Р.Е. Алексеева, AuthorID: 531326, ORCID: 0000-0002-3644-5234, iamuser@inbox.ru, Россия, Нижний Новгород, ш. Казанское, 21, 170.*

*Смирнов Александр Вадимович. Магистр кафедры Вычислительные системы и технологии, старший инженер по разработке графических систем А/О Интел АО, ORCID: 0000-0002-4852-8184, mrroman152@gmail.com, Россия, Нижний Новгород, ш. Казанское, 21, 170.*

*Баринов Роман Олегович. Магистр кафедры Вычислительные системы и технологии, ИРИТ, НГТУ им. Р.Е. Алексеева, инженер ООО «СКБ «Связь и локация», ORCID: 0000-0002-0269-3205, barinovr@list.ru, Россия, Нижний Новгород, ш. Казанское, 21, 170.*

*Поляков Игорь Владимирович. Ассистент, аспирант кафедры Вычислительные системы и технологии, ИРИТ, НГТУ им. Р.Е. Алексеева, AuthorID: 879102, ORCID: 0000-0002-1492-9350, polyakovigor92@gmail.com, Россия, Нижний Новгород, ш. Казанское, 21, 170.*

*Голубенко Владимир Андреевич. Магистр кафедры Вычислительные системы и технологии, ИРИТ, НГТУ им. Р.Е. Алексеева, практикант-разработчик ПО А/О Интел АО, ORCID: 0000-0002-4683-3249, fullmoon-shrine@gmail.com, Россия, Нижний Новгород, ш. Казанское, 21, 170.*

*Кузнецов Георгий Дмитриевич. Магистр кафедры Вычислительные системы и технологии, ИРИТ, НГТУ им. Р.Е. Алексеева, инженер ООО «СКБ «Связь и локация», ORCID: 0000-0001-5564-045X, gd.smith@yandex.ru, Россия, Нижний Новгород, ш. Казанское, 21, 170.*

---

UDC 004.932.72

DOI:10.38028/ESI.2022.28.4.016

## **Object recognition system using theory of active perception**

**Vasily E. Gai, Alexander V. Smirnov, Roman O. Barinov, Igor V. Polyakov,  
Vladimir A. Golubenko, Georgy D. Kuznetsov**

Nizhny Novgorod State Technical University,  
Russia, Nizhny Novgorod, *barinovr@list.ru*

**Abstract.** In this paper we propose a method for detecting an object in an image based on a global feature description. An information model is described and implementation options for each of the stages of the image object detection task are proposed. For the image preprocessing stage, the available options, including normalization, calculation of brightness function and application of Gaussian filter are given. The stage of forming a global feature description of an object is based on active perception theory (U-transformation). Object localization is performed based on the minimum Euclidean distance from the detected object to the reference objects from the database. Images from the Russian Traffic Sign Dataset database and their modified copies (images with superimposed noise, images with rotation of the objects searched for) were used for testing. When

analyzing the test results, the parameters that give the highest accuracy for the proposed method of object detection in the image have been selected. In the presence of noise in the image, the localization accuracy of the proposed method was more than 70%. The proposed image object detection method is robust to rotation of the objects being searched. The resulting accuracy of about 94-96%, when compared with the accuracy of existing methods, showed that under normal conditions the developed method works similarly to existing methods.

**Keywords:** image processing, object detection, theory of active perception

## References

1. Oliva A., Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope, *Int. J. Comput. Vis.*, 2001, vol. 42, no. 3, pp. 145-175.
2. Dalal N., Triggs B. Histograms of oriented gradients for human detection. In *CVPR*, 2005.
3. Watanabe T., Ito S., Yokoi K. "Co-occurrence histograms of oriented gradients for pedestrian detection". *PSIVT*, 2009.
4. Hu M.K. Visual Pattern recognition by moment invariants, *IRE trans. Info. Theory*, 1962, vol. IT-8, pp.179–187.
5. Abramov N.S., Fralenko V.P. Opredeleniye rasstoyaniy na osnove sistemy tekhnicheskogo zreniya i metoda invariantnykh momentov [Determination of distances based on the technical vision system and the method of invariant moments]. *Informatsionnyye tekhnologii i vychislitel'nyye sistemy [Information technologies and computational systems]*, 2012, no. 4, pp. 32-39.
6. Rodzhers D., Adams Dzh., *Matematicheskiye osnovy mashinnoy grafiki [Mathematical Foundations of Computer Graphics]*. M.: Mashinostroyeniye [Mechanical engineering], 1980, 240 c.
7. Utrobin V. A. Elementy teorii aktivnogo vospriyatiya izobrazheniy [Elements of the theory of active image perception]. *Izvestiya Nizhegorodskogo gosudarstvennogo tekhnicheskogo universiteta [Proceedings of Nizhny Novgorod State Technical University]*, 2001, 64 p.
8. Kondratiev V.V., Zhevnerchuk D.V. Application of methods of self-organization theory to problems of profiling and configuring computational systems. *Doklady Mathematics*, Pleiades Publishing, Ltd., 2014, vol. 90, no. 3, pp. 788-190, ISSN 1064-5624.
9. Haddad R.A., Akansu A.N. A class of fast gaussian binomial filters for speech and image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1991, vol. 39, pp 723-727.
10. Ramer Urs. An iterative procedure for the polygonal approximation of plane curves. *Computer graphics and image processing*, 1972, no. 1(3), pp. 244-256, DOI: 10.1016/S0146-664X(72)80017-0.
11. Shakhuro V., Konushin A. Rossiyskaya baza izobrazheniy avtodorozhnykh znakov [Russian database of images of road signs]. *Komp'yuternaya optika [Computer Optics]*, 2016.

**Vasily Evgenievich Gai.** Engineer, Associate Professor, NSTU, AuthorID: 531326, ORCID: 0000-0002-3644-5234, iamuser@inbox.ru, Russia, Nizhny Novgorod, sh. Kazanskoe, 21, 170.

**Alexander Vadimovich Smirnov.** Master of the Department of Computing Systems and Technologies, Senior Engineer for the Development of Graphic Systems of A/O Intel JSC, ORCID: 0000-0002-4852-8184, mrroman152@gmail.com, Russia, Nizhny Novgorod, sh. Kazanskoe, 21, 170.

**Roman Olegovich Barinov.** Master of the Department of Computing Systems and Technologies, IRIT, NSTU, engineer of SKB Svyaz and Location LLC, ORCID: 0000-0002-0269-3205, barinovr@list.ru, Russia, Nizhny Novgorod, sh. Kazanskoe, 21, 170.

**Igor Vladimirovich Polyakov.** Assistant, postgraduate student of the Department of Computing Systems and Technologies, IRIT, NSTU, AuthorID: 879102, ORCID: 0000-0002-1492-9350, polyakovigor92@gmail.com, Russia, Nizhny Novgorod, sh. Kazanskoe, 21, 170.

**Vladimir Andreevich Golubenko.** Master of the Department of Computing Systems and Technologies, IRIT, NSTU, Intern-Software Developer at JSC Intel JSC, ORCID: 0000-0002-4683-3249, fullmoonshrine@gmail.com, Russia, Nizhny Novgorod, sh. Kazanskoe, 21, 170.

**Georgy Dmitrievich Kuznetsov.** Master of the Department of Computing Systems and Technologies, IRIT, NSTU, engineer of SKB Svyaz and Location LLC, ORCID: 0000-0001-5564-045X, gd.smith@yandex.ru, Russia, Nizhny Novgorod, sh. Kazanskoe, 21, 170.

Статья поступила в редакцию 14.07.2022; одобрена после рецензирования 14.09.2022; принята к публикации 19.09.2022.

The article was submitted 07/14/2022; approved after reviewing 09/14/2022; accepted for publication 09/19/2022.