

УДК 004.932.72

DOI:10.38028/ESI.2023.29.1.003

## Модель и алгоритмы обнаружения объектов на изображении с использованием локального признакового описания

Гай Василий Евгеньевич, Домнина Наталья Александровна,

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*

**Аннотация.** Предлагается метод обнаружения объекта на изображении на основе теории активного восприятия. Теория активного восприятия предоставляет возможность для реализации этапов предварительной обработки и формирования признакового описания. Признаковое описание областей изображения, состоящее из спектральных коэффициентов  $U$ -преобразования, позволяет выявлять области с перепадами яркости. Для принятия решения, является ли рассматриваемая область ключевой, то есть, содержит ли область перепады яркости, соответствующие контуру объекта, вводится коэффициент, который в совокупности с максимальным значением стандартного отклонения дает порог, уникальный для каждого изображения. Для достижения инвариантности к повороту и масштабу, изображение-шаблон подвергается масштабированию и вращению. Локализация целевого объекта осуществляется методом  $k$ -means. Для тестирования использовались изображения из базы Amsterdam library of object images, их изменённые копии (несколько целевых объектов на исследуемом изображении, изображения с наложенным шумом) и реальные изображения. Для каждого типа изображений было проведено сравнение результатов предложенного метода с результатами метода scale-invariant feature transform. При анализе результатов тестирования были предложены параметры, дающие наибольшую точность предложенного метода обнаружения объекта на изображении. При наличии шума на изображении, точность предложенного метода в сравнении с точностью существующего метода оказалась больше на 30 %. При наличии на тестовом изображении более одного искомого объекта предложенный метод с точностью порядка 96 % способен определить все целевые объекты.

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

**Цитирование:** Гай В.Е. Модель и алгоритмы обнаружения объектов на изображении с использованием локального признакового описания / В.Е. Гай, Н.А. Домнина, Р.О. Баринов, И.В. Поляков, В.А. Голубенко, Г.Д. Кузнецов // Информационные и математические технологии в науке и управлении. – 2023. – № 1(29). – С. 33-43. – DOI:10.38028/ESI.2023.29.1.003.

**Введение.** В настоящее время наблюдается быстрый рост объемов цифровых графических данных. Каждую секунду в сеть Интернет попадает множество фото и видео с разных концов света. Посещая музей, галерею, другой город или другую страну люди хотят запечатлеть каждый значимый момент этого события. В связи с этим число алгоритмов, применимых к фото- и видеосъемке, с каждым днем все растет. Уже никого не удивляет, что при создании фото на экране отображается возраст и пол человека, а при создании группового фото выделяются лица всех людей. Гораздо меньше внимания уделено анализу достопримечательностей и произведений искусства.

Чтобы при наведении камеры отображалось название памятника или скульптуры, нужно разработать быстрый алгоритм для использования его в реальном времени, а также обеспечить устойчивость к повороту, шуму, смене освещения и масштаба. Эффективность этого алгоритма в наибольшей мере зависит от того, как изображение будет представлено для анализа. Здесь очень важно правильно выбрать так называемые «ключевые» точки и способ их описания.

### 1. Обзор существующих методов обнаружения объектов на изображении.

Существующие методы обнаружения объектов на изображении преимущественно включают в себя три этапа:

1. Предварительная обработка изображения – преобразование изображения к виду, подходящему для вычисления признаков.
2. Формирование признакового описания – процесс создания итогового описания изображения.
3. Принятие решения о присутствии объекта на изображении, а также его локализация – процесс сопоставления признаков двух изображений.

Структура системы обнаружения объектов представлена на рисунке 1.



**Рис. 1.** Структура системы обнаружения объектов

Предварительная обработка изображения, как правило, производится для избавления от дефектов изображения и повышения точности распознавания.

Для снижения уровня шума на изображениях разработано много методов и алгоритмов. Среди них можно выделить методы адаптивного сглаживания [1], анизотропной диффузии [2] и медианной фильтрации [1]. Последний обладает высоким быстродействием и позволяет снизить уровень помех на изображениях, при его работе не происходит усреднения по яркости, контуры объектов на изображении не размываются.

Метод формирования описания изображения на основе локальных бинарных шаблонов представлен в [3]. Он основан на вычислении знака разности между яркостями соседних отсчётов. В методе SIFT [4] признаковое описание вычисляется по ключевым точкам. В качестве них выступают так называемые «пятна» – это такие области, которые ярче или темнее, чем их окружение. Такой выбор был сделан по той причине, что на маленькие области меньшее влияние оказывают эффекты искажений, а некоторые изменения могут вообще не повлиять на дескриптор. Построение дескрипторов в алгоритме SURF [5] основывается на тех же принципах, что и SIFT. В методе BRIEF [6] для этой цели вычисляется расстояния Хэмминга [7] между двумя последовательностями битов. Если дескрипторы представлены не в виде бинарных или символьных строк, то в качестве метрики может использовать Евклидово расстояние. Оно применимо для признаков, описываемых количественными переменными.

Вычисление вышеописанных метрик может быть финальным этапом в сопоставлении признаков. В этом случае есть два подхода:

1. Каждому дескриптору изображения-шаблона ставится в соответствие некоторое количество дескрипторов тестируемого изображения. Совпадение дескрипторов определяется порогом.
2. Каждому дескриптору изображения-шаблона ставится в соответствие один наиболее подходящий дескриптор тестируемого изображения.

Другой вариант, например, алгоритм RANSAC [8], может содержать дополнительные шаги для локализации объекта. Если объект переместился на сцене или снят с другого

ракурса, то можно выявить особенности, не относящиеся к объекту и тем самым уменьшить количество ложно определенных связей.

После того, как все признаки сопоставлены и, возможно, применены дополнительные методы для достижения наилучшего результата, остается локализовать объект.

Большинство существующих алгоритмов сегментации можно разделить на два класса [9]:

- интерактивные – использующие пользовательские подсказки;
- автоматические – не требующие участия пользователя;
- автоматические алгоритмы сегментации, которые, в свою очередь, также делятся на два класса:
  - выделение областей изображения с заданными свойствами, уникальными для конкретной предметной области (mean shift [10, 11], k-means [12]);
  - разбиение изображения на однородные области.

**2. Реализация системы обнаружения объекта на изображении на основе теории активного восприятия.** Теория активного восприятия [13, 14] позволяет реализовать как этап предварительной обработки изображения, так и этап формирования признакового описания.

Изображение представляется в виде множества областей, которые получены из исходного с помощью скользящего окна:

$$P = \{P_{ij}\}, i \in X, j \in Y,$$

где  $P_{ij}$  – область изображения размером  $n \times n$  пикселей;  $i, j$  – координаты левого верхнего угла области  $P_{ij}$ .

Для изображения-шаблона на этом этапе выполняются еще несколько шагов. Чтобы добиться инвариантности к повороту и масштабу, после представления изображения в градациях серого, выполняется его масштабирование с коэффициентами  $s$ , а также вращение на углы  $\pm\alpha$ .

Признаковое описание для каждой области изображения  $P_{ij}$  можно представить в виде следующей формулы:

$$D_{ij} = \{d_k\}, k = \overline{1, L},$$

где  $d_k$  –  $k$ -й коэффициент спектрального представления области изображения;  $L$  – количество фильтров, применяемых при вычислении  $U$ -преобразования.

Исходя из этого признаковое описание изображения в целом можно представить в виде набора признаковых описаний всех его областей:

$$D = \{D_{ij}\}, i \in X, j \in Y.$$

Для оптимизации системы в итоговые дескрипторы изображения заносятся только те признаки, которые соответствуют ключевым областям. Будем считать область ключевой, если в ней есть перепад яркости, который соответствует контуру объекта.

Признаковое описание областей изображения, состоящее из спектральных коэффициентов  $U$ -преобразования, позволяет выявлять области с перепадами яркости. Для каждой области, по которой были вычислены спектральные коэффициенты, вычисляется их среднеквадратичное отклонение (СКО). Исходя из этого, получаем формулу для вычисления СКО спектральных коэффициентов области  $P_{ij}$ :

$$s_{ij} = \sqrt{\frac{1}{n-1} \sum_1^n (d_i - d_{cp})^2},$$

где  $n = 15$  – количество учитываемых коэффициентов спектрального представления;  $d_i$  –  $i$ -й спектральный коэффициент;  $d_{cp}$  – среднее арифметическое спектральных коэффициентов рассматриваемой области без учета 0-го коэффициента.

Для принятия решения, является ли рассматриваемая область ключевой, введем коэффициент, который в применении с максимальным значением СКО будет давать некий порог, уникальный для каждого изображения:

$$s_{ij} > kS_{\max},$$

где  $k$  – задаваемый коэффициент отбора точек,  $S_{\max}$  – максимальное значение СКО по всем областям изображения.

Увеличение коэффициента  $k$  задает больший порог, что приводит к уменьшению количества точек, принимаемых за ключевые. Чем ниже его значение, тем меньший перепад яркости области нужен для того, чтобы считать область ключевой.

Последний этап служит для сопоставления дескрипторов исследуемого изображения и изображения-шаблона.

Для каждой точки изображения-шаблона вычисляется расстояние до каждой из точек исследуемого изображения. В результате этого получаем вектор, который содержит 5 значений: вычисленное расстояние, координаты точки первого изображения и координаты точки второго изображения.

После того, как точки были сопоставлены, необходимо локализовать объект. Для этой цели предлагается использовать алгоритм k-means. Поскольку алгоритм требует указать количество сегментов, на которое будет поделено изображение, один его вызов с некоторым заданным числом может быть не всегда эффективен. Для решения этой проблемы было решено поступить следующим образом. Первый вызов k-means осуществляется с минимальным числом сегментов – 2. После этого используется вспомогательный метод выборочного поиска [14], который из полученных сегментов создаст набор областей, которые в дальнейшем будут проверяться на наличие объекта с помощью уже сопоставленных нами точек. Данный метод осуществляет выборочный поиск, используя 4 показателя сходства, основанные на совместимости цвета, текстуры, размера и формы. Выборочный поиск начинается с чрезмерного сегментирования изображения на основе интенсивности пикселей с использованием метода сегментации на основе графиков [16].

Таким образом, с помощью данного метода формируется набор областей для дальнейшего принятия решения о наличии на одной или нескольких найденных областях искомого объекта.

Следующим шагом является поиск области, содержащей объект. Для этого, с использованием координат найденных областей и координат сопоставленных точек, подсчитывается количество таких точек в каждой из областей. Каждая область, содержащая не менее  $m$  совпадений ключевых точек, считается за искомым объектом.

**3. Вычислительный эксперимент.** Для проведения эксперимента использовалась база изображений 1000 предметов, созданная в Амстердамском университете – ALOI [17].

Для каждого объекта в этой базе есть одно или несколько следующих видов изменений: угол обзора, направление освещения, цвет освещения.

Для проведения тестирования были выбраны 100 объектов с разными типами изменений. В дополнение к этому, с использованием этой базы были получены дополнительные изображения для проверки устойчивости алгоритма к шуму, масштабу и присутствию на исследуемом изображении нескольких искомым объектов. Для создания изображений разного масштаба использовались следующие коэффициенты  $S$ : 0,9; 0,8; 0,7; 0,6; 0,5; 0,4; 0,3; 0,2.

Для искажения изображений использовался шум «соль-перец». Были использованы следующие значения вероятностей для создания такого шума: 0,02; 0,04; 0,06; 0,1; 0,2; 0,4.

Последней частью набора данных для тестирования являются реальные изображения. Основная часть – это фотографии памятников, остальные – фотографии обычных предметов, но, в отличие от ALOI, это реальные фотографии с присутствием некоторого фона.

При поиске области, содержащей объект, количество совпавших ключевых точек  $m$  задаётся равным 30 для всех случаев.

В итоге, получается база изображений 435 предметов, у каждого предмета есть шаблоны и изображения для поиска с разными типами изменений объекта.

**3.1. Анализ результатов эксперимента.** Разделим результаты в зависимости от вида тестового изображения:

1. Изображения из базы ALOI.
2. Изображения, созданные с использованием базы ALOI (несколько искомым объектов на исследуемом изображении, изображения с наложенным шумом).
3. Реальные изображения.

В таблицах будут представлены результаты для разных значений параметров, которые используются в разрабатываемой системе. Для каждого типа входных изображений будут также приведены результаты метода SIFT для сравнения точности и времени работы.

Приведем описание параметров:

- $k_{kp}$  – коэффициент отбора ключевых точек; чем меньше значение, тем больше на изображении будет определено ключевых точек
- $k_m$  – коэффициент отбора дескрипторов; чем больше значение, тем больше будет определено совпадений
- $n$  – размер скользящего окна

Для каждой комбинации параметров будет указана точность локализации, а также затрачиваемое на локализацию время.

В таблице 1 приведены результаты тестирования предложенного алгоритма на устойчивость к изменению угла обзора. Данный блок содержит 60 изображений отдельных объектов на однородном фоне. У тестового изображения изменен угол обзора  $\alpha$  от -30 до +30 с шагом в 10 градусов.

**Таблица 1.** Устойчивость к изменению угла обзора

$k_{kp}$	$k_m$	$n$	Точность
0,5	1,6	16	96,67%
0,6	1,8	16	96,67%

В таблице 2 приведены результаты тестирования предложенного алгоритма на устойчивость к изменению направления освещения. Данный блок содержит 50 изображений отдельных объектов на однородном фоне, у тестового изображения изменено направление освещения.

**Таблица 2.** Устойчивость к изменению направления освещения

$k_{kp}$	$k_m$	$n$	Точность
0,5	1,6	16	100,00%
		32	100,00%
	1,8	16	100,00%
0,6	1,6	16	100,00%
	1,8	16	100,00%

Из результатов можно сделать вывод, что метод полностью устойчив к изменению освещению. Результат, полученный с использованием SIFT, также дает точность 100%.

Результаты тестирования предложенного алгоритма на изображениях с изменением цвета освещения представлены в таблице 3. Данный блок содержит 45 изображений отдельных объектов на однородном фоне, у тестового изображения изменен цвет освещения.

**Таблица 3.** Устойчивость к изменению цвета освещения

$k_{кр}$	$k_m$	$n$	Точность
0,5	1,4	16	100,00%
		16	100,00%
	1,6	32	100,00%
		16	100,00%
0,6	1,8	16	100,00%
		32	100,00%
	1,6	16	100,00%
		16	100,00%
2,0	16	100,00%	

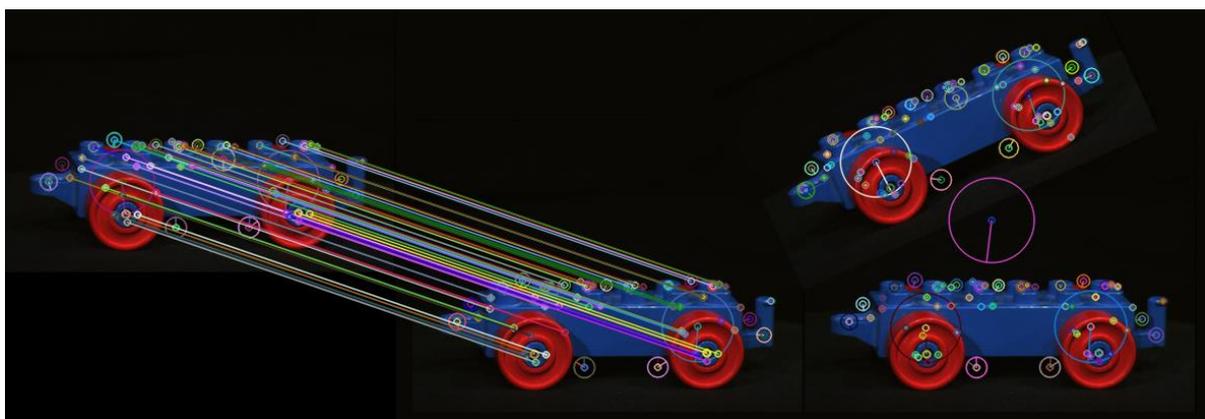
Из результатов можно сделать вывод, что метод полностью устойчив к изменению цвета освещения объекта. Результат, полученный с использованием SIFT, также дает точность 100%.

В таблице 4 представлены результаты тестирования предложенного алгоритма на способность к обнаружению нескольких объектов на изображении (отдельные объекты на однородном фоне, тестовое изображение имеет 2 и более искомым объекта, которые могут быть повернуты, наклонены, а также иметь измененный угол обзора или масштаб). Данный блок содержит 30 изображений.

**Таблица 4.** Несколько объектов на изображении

$k_{кр}$	$k_m$	$n$	Точность (минимум 1 объект)	Точность (все объекты)
0,5	1,6	16	100,00%	93,33%
	1,8	16	100,00%	93,33%
0,6	1,4	16	100,00%	93,33%
	1,6	16	100,00%	93,33%

Метод SIFT определяет один объект с точностью 100.00 %, но он не может определить 2 и более объекта, поэтому здесь его точность – 0.00 %. Пример работы метода SIFT с несколькими искомыми объектами представлен на рисунке 2.



**Рис. 2.** Результат работы метода SIFT с несколькими объектами

Метод SIFT корректно нашел ключевые точки на всех 3 объектах, но на этапе сопоставления дескрипторов, он сопоставил только максимально похожие. Для сравнения на рисунке 3 показан результат работы разрабатываемого метода на том же тестовом изображении.

В таблице 5 приведены результаты тестирования предложенного алгоритма на способность к обнаружению нескольких объектов на изображении с шумом. Данный блок содержит 30 изображений.



Рис. 3. Результат работы предложенного метода с несколькими объектами

Таблица 5. Несколько объектов на изображении с шумом

$k_{кр}$	$k_m$	$n$	Точность (минимум 1 объект)	Точность (все объекты)
0,5	1,8	32	96,67%	90,00%
0,6	1,6	16	96,67%	90,00%

Метод SIFT оказался очень слабо устойчив к шуму, он определяет один объект с точностью 38.33%, он также не может определить 2 и более объектов, поэтому здесь его точность – 0.00%. На рисунке 4 представлен результат работы разрабатываемого метода при наличии шума на изображении.

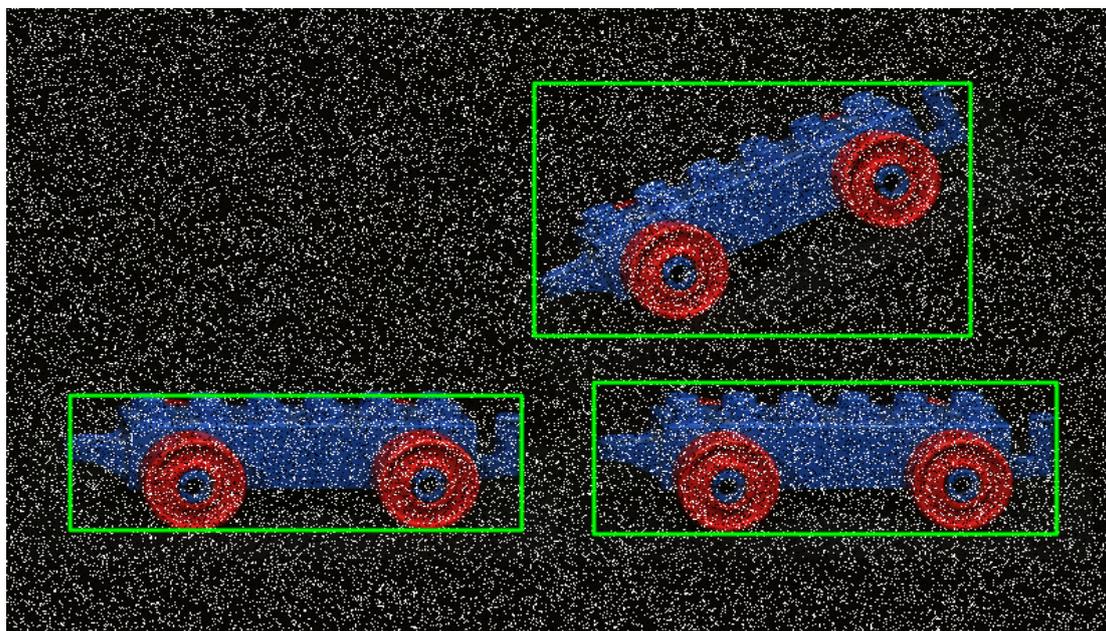


Рис. 4. Результат работы предложенного метода при наличии шума на изображении

В дополнение к этому виду тестирования рассмотрим еще 80 изображений, которые соответствуют 10 объектам, каждый из которых имеет следующие вероятности наложенного шума «соль-перец»: 0,02; 0,04; 0,06; 0,1; 0,2; 0,4; 0,5; 0,6. Из результатов тестов следует, что вплоть до вероятности шума, равной 10%, точность определения объекта составляет 100%, вероятности 20% и 40% дают точность 90%, что соответствует 9 найденным объектам из 10. Начиная с вероятности 50% точность начинает резко падать и показывает 60% при вероятности шума 50% и 40% точности при вероятности шума 60%.

Таким образом, можно сделать вывод, что метод SIFT нельзя применять в ситуациях, когда есть даже небольшой шум. При вероятности шума 2% он уже показывает точность 60%, при вероятности 4% точность снижается до 50% и с повышением уровня шума только падает. Вероятно, причина этого – повышение контрастности на изображении на этапе предварительной обработки данных и отсутствие применения алгоритмов, понижающих шум.

Рассмотрим еще один дополнительный блок, который включает в себя 80 изображений, соответствующих 10 объектам, каждый из которых имеет следующие коэффициенты масштабирования  $s$ : 0,9; 0,8; 0,7; 0,6; 0,5; 0,4; 0,3; 0,2. Результаты приведены в таблице 6.

**Таблица 6.** Результаты дополнительного тестирования

Масштаб	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2
Точность	100%	100%	100%	100%	100%	100%	90%	50%

Исходя из полученных данных, можно сделать вывод, что предложенный алгоритм устойчив к изменению масштаба. При коэффициенте масштабирования 0,2 метод SIFT корректно определил 8 объектов из 10, в то время как разрабатываемый справился с 5 объектами.

Результаты тестирования предложенного алгоритма на реальных изображениях представлены в таблице 7. Данный блок содержит 40 изображений, на которых могут быть изменены угол обзора, освещение, масштаб, а также присутствовать шум.

**Таблица 7.** Реальные изображения

$k_{кр}$	$k_m$	$n$	Точность
0,5	1,8	16	95,00%
0,6	1,4	32	95,00%
	1,8	16	95,00%

Результат, полученный с использованием метода SIFT, дает точность 85,0%. Если рассматривать результаты метода SIFT без зашумленных изображений, то его точность равна 96,67%.

Таким образом, для каждого конкретного случая были выявлены параметры, дающие наиболее высокую точность. Суммируя полученные результаты, можно определить параметры, которые дают лучшие результаты в большинстве случаев. Такими параметрами являются:  $k_{кр} - 0,5$ ,  $k_m - 1,8$ ,  $n - 16$ .

Среднее время обработки изображения – 0,24 секунды.

Анализируя результаты тестов, можно сделать вывод о том, метод SIFT работает лучше разрабатываемого в нормальных условиях: он дает 100% точность при изменении угла поворота, направления и цвета освещения для изображений из тестовой базы и 96,67% точность для реальных изображений. Однако, на зашумленных изображениях его точность снижается до 63,33%. Также метод SIFT не может работать с несколькими объектами на изображении. При этом время работы метода SIFT вдвое меньше времени разрабатываемого метода (0,24 секунд против 0,10 секунд).

**Заключение.** Проведённые эксперименты и сравнительный анализ показали, что точность работы системы в нормальных условиях оказалась сопоставимой с точностью существующих методов, однако при наличии шума разработанный метод дает намного большую точность (+ 30%).

Ряд существующих методов (SIFT, SURF) на этапе сопоставления дескрипторов исследуемого изображения и изображения-шаблона находят только одну наиболее подходящую точку, вследствие чего такие методы не могут обнаружить 2 и более объектов, если они присутствуют на изображении. Сопоставление точки «один – ко всем» занимает

немного больше времени, нежели останавливать поиск на первой подходящей, но такой подход значительно расширяет функциональность метода, позволяя находить все присутствующие на изображении объекты с точностью 96,67%.

### Список источников

1. Писаревский А.Н. Системы технического зрения (принципиальные основы, аппаратное и математическое обеспечение) / А.Н. Писаревский. – Л.: Машиностроение, 1988. – 424 с.
2. Анисимов Б.В. Распознавание и цифровая обработка изображений: Учеб. Пособие / Б.В. Анисимов, В.Д. Курганов, В.К. Злобин. – М.: Высш. школа, 1983. – 295 с.
3. Sharma G., ul Hussain, Jurie F. Local higher-order statistics (LHS) for texture categorization and facial analysis. Computer Vision – ECCV 2012, 12th European Conference on Computer Vision, Florence, Italy, 2012, Proceedings, Part VII, Berlin, Heidelberg, Springer Berlin Heidelberg, 2012, pp. 1-12.
4. Oji R., Tajeripour F. Full object boundary detection by applying scale invariant features in a region merging segmentation algorithm. International journal of artificial intelligence & applications (IJAA), 2012, vol. 3, no. 5, pp. 41-50.
5. Bay H., Ess A., Tuytelaars T., Van Gool L. SURF: Speeded up robust features. Computer vision image understanding, 2008, vol. 110, no. 3, pp. 346-359.
6. Chen J. RLBP: Robust Local Binary Pattern. Proceedings of the British Machine Vision Conference, BMVA Press, 2013.
7. Minsky M., Papert S. Perceptrons. MIT Press, Cambridge, MA, 1969.
8. Martin A. Fischler, Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. Of the ACM, vol. 24, pp. 381-395.
9. Шапиро Л. Компьютерное зрение / Л. Шапиро, Дж. Стокман. – М.: Бином: Лаборатория знаний, 2006. – 752 с.
10. Comaniciu D., Meer P. Mean shift: A robust approach towards feature space analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, 24: 603–619, 2002.
11. Kim E., Wang W., Li H., Huang X. A parallel annealing method for automatic color cervigram image segmentation. In: Medical Image computing and computer assisted intervention, MICCAI-GRID 2009 HPC Workshop, 2009.
12. Вежневцев А., Баринаова О. Методы сегментации изображений: автоматическая сегментация / А. Вежневцев, О. Баринаова // Компьютерная графика и мультимедиа, 2006. – № 4.
13. Утробин В.А. Элементы теории активного восприятия изображений / В.А. Утробин // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева, 2010. – Т. 81. – №2. – С. 61-69.
14. Кондратьев В.В. Применение методов теории самоорганизации к задачам профилирования и конфигурирования вычислительных систем / В.В. Кондратьев, Д.В. Жевнерчук // Доклады математики. – Издательство Плейды, 2014. – Т. 90. – №. 3. – С. 788-790. – ISSN 1064-5624.
15. Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, Arnold WM Smeulders. Selective search for object recognition. International journal of computer vision, 104(2):154–171, 2013.
16. Felzenszwalb P., Huttenlocher D., Efficient graph-based image segmentation. International Journal of Computer Vision, 2004, vol. 59, no. 2.
17. Geusebroek J.M. Burghouts G.J., Smeulders A.W.M, The Amsterdam library of object images. Int. J. Comput. Vision, 2005, 61(1), pp. 103-112.

**Гай Василий Евгеньевич.** Доцент НГТУ им. Р.Е. Алексеева, AuthorID: 531326, ORCID: 0000-0002-3644-5234, iamuser@inbox.ru.

**Домнина Наталья Александровна.** Старший инженер по разработке программного обеспечения в области компьютерной графики А/О Интел АО, ORCID: 0000-0002-3817-401X, zhestckova.natali@yandex.ru.

**Баринов Роман Олегович.** Инженер ООО «СКБ «Связь и локация», ORCID: 0000-0002-0269-3205, barinovr@list.ru.

**Поляков Игорь Владимирович.** Ассистент НГТУ им. Р.Е. Алексеева, AuthorID: 879102, ORCID: 0000-0002-1492-9350, polyakovigor92@gmail.com.

**Голубенко Владимир Андреевич.** Практикант-разработчик ПО А/О Интел АО, ORCID: 0000-0002-4683-3249, fullmoonshrine@gmail.com.

**Кузнецов Георгий Дмитриевич.** Инженер, ООО «СКБ «Связь и локация», ORCID: 0000-0001-5564-045X, gd.smith@yandex.ru.

UDC 004.932.72

DOI:10.38028/ESI.2023.29.1.003

## Model and algorithms for detecting objects in an image using a local feature description

Vasily E. Gai, Natalia A. Domnina, Roman O. Barinov, Igor V. Polyakov, Vladimir A. Golubenko, Georgy D. Kuznetsov

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

**Abstract.** The paper proposes an object detection method based on the theory of active perception. The theory of active perception provides an opportunity for implementation of pre-processing and feature description stages. The region feature descriptors consisting of spectral coefficients of U-transformation allows one to identify regions with brightness variations. To designate a region as key, that is containing brightness differences corresponding to the object contour, a coefficient is introduced, which in combination with maximum RMS results in a unique threshold for each image. To achieve invariance to rotation and scale, the template image is subjected to scale and rotation transformations. Localization of the target object is done using the k-means method. For testing, images from the ALOI database, as well as their altered copies (multiple target objects in the image under study, images with superimposed noise) were used. For each image type, the results of the proposed method were compared with the scale-invariant feature transform method. The parameters resulting in the highest accuracy for the proposed object detection method were proposed for analysis. On noisy images, accuracy of the proposed method increased by 30% compared to the existing method. With more than one target object present on the test image, the proposed method was able to detect all target objects with an accuracy of about 96%.

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

### References

1. Pisarevsky A.N. Sistemy tekhnicheskogo zreniya (printsipial'nyye osnovy, apparatnoye i matematicheskoye obespecheniye) [Technical Vision Systems (Fundamental Principles, Hardware and Mathematical Support)]. Mashinostroenie, 1988. 424 p.
2. Anisimov B.V., Kurganov V.D., Zlobin V.K. Raspoznavaniye i tsifrovaya obrabotka izobrazheniy, Ucheb. Posobiye [Recognition and digital image processing, Textbook]. Moscow: High School [Higher. School], 1983, 295 p.
3. Sharma G., ul Hussain, Jurie F. Local higher-order statistics (LHS) for texture categorization and facial analysis. Computer Vision – ECCV 2012, 12th European Conference on Computer Vision, Florence, Italy, 2012, Proceedings, Part VII, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1-12.
4. Oji R., Tajeripour F. Full object boundary detection by applying scale invariant features in a region merging segmentation algorithm. International journal of artificial intelligence & applications (IJAIA), 2012, vol. 3, no. 5, pp. 41-50.
5. Bay H., Ess A., Tuytelaars T., Van Gool L. SURF: Speeded up robust features. Computer vision image understanding, 2008, vol. 110, no. 3, pp. 346-359.
6. Chen J. RLBP: Robust Local Binary Pattern. Proceedings of the British Machine Vision Conference. BMVA Press, 2013.
7. Minsky M., Papert S. Perceptrons. MIT Press, Cambridge, MA, 1969.
8. Martin A. Fischler, Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. Of the ACM, vol. 24, pp. 381-395.
9. Shapiro L., Stockman J. Komp'yuternoye zreniye [Computer Vision]. Moscow: Binom, Laboratoriya znaniy [Laboratory of Knowledge], 2006, 752 p.
10. Comaniciu D., Meer P. Mean shift: A robust approach towards feature space analysis. IEEE Trans. Pattern Analysis and Machine Intelligence, 24: 603–619, 2002.
11. Kim E., Wang W., Li H., Huang X. A parallel annealing method for automatic color cervigram image segmentation. In: Medical Image computing and computer assisted intervention, MICCAI-GRID 2009 HPC Workshop, 2009.
12. Vezhnevets A., Barinova O. Metody segmentatsii izobrazheniy: avtomaticheskaya seg-mentatsiya [Methods of Image Segmentation: Automatic Segmentation]. Komp'yuternaya grafika i mul'timedia [Computer graphics and multimedia], 2006, no. 4.

13. Utrobin V.A. Elementy teorii aktivnogo vospriyatiya izobrazheniy [Elements of the theory of active image perception]. Trudy Nizhegorodskogo gosudartsvennogo tekhnicheskogo universiteta im. R.Ye. Alekseyeva [Proceedings of Nizhny Novgorod State Technical University named after R.E. Alekseev], 2010, vol. 81, no.2, pp. 61-69.
14. Kondratiev V.V., Zhevnerchuk D.V. Primeneniye metodov teorii samoorganizatsii k zadacham profilirovaniya i konfigurirovaniya vychislitel'nykh sistem [Application of Methods of Self-Organization Theory to Problems of Profiling and Configuring Computational Systems]. Doklady Mathematics [Reports of Mathematics], Izdatel'stvo Pleyady [Reports of Mathematics], 2014, vol. 90, no. 3, pp. 788-190, ISSN 1064-5624.
15. Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, Arnold WM Smeulders. Selective search for object recognition. International journal of computer vision, 104(2):154–171, 2013.
16. Felzenszwalb P., Huttenlocher D., Efficient graph-based image segmentation. International Journal of Computer Vision, 2004, vol. 59, no. 2.
17. Geusebroek J.M. Burghouts G.J., Smeulders A.W.M, The Amsterdam library of object images. Int. J. Comput. Vision, 2005, 61(1), pp. 103-112.

**Gai Vasily Evgenievich.** Associate Professor at NSTU R.E. Alekseeva, AuthorID: 531326, ORCID: 0000-0002-3644-5234, iamuser@inbox.ru.

**Domnina Natalia Alexandrovna.** Senior engineer for software development in the field of computer graphics of Intel A/O, ORCID: 0000-0002-3817-401X, zhestckova.natali@yandex.ru.

**Barinov Roman Olegovich.** Engineer of SKB «Communication and location», ORCID: 0000-0002-0269-3205, barinovr@list.ru.

**Polyakov Igor Vladimirovich.** Assistant at NSTU R.E. Alekseeva, AuthorID: 879102, ORCID: 0000-0002-1492-9350, polyakovigor92@gmail.com.

**Golubenko Vladimir Andreevich.** Intern-software developer of Intel A/O, ORCID: 0000-0002-4683-3249, fullmoonshrine@gmail.com.

**Kuznetsov Georgy Dmitrievich.** Engineer of SKB «Communication and location», ORCID: 0000-0001-5564-045X, gd.smith@yandex.ru.

Статья поступила в редакцию 14.07.2022; одобрена после рецензирования 07.09.2022; принята к публикации 15.02.2023.

The article was submitted 07/14/2022; approved after reviewing 09/07/2022; accepted for publication 02/15/2023.