электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51038.

УДК 004.932

Метод выделения похожих изображений на основе применения SIFT дескрипторов

***Савонин А.И.**, студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Программное обеспечение ЭВМ и информационные технологии»

Научный руководитель: Рогозин О.В., к.т.н, доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана
irudakov@bmstu.ru

Введение

Повсеместное распространение технологий и доступа к Интернету привели к увеличению объёмов информации, большая часть которой представлена в текстовой и графической форме. Но далеко не вся информация является уникальной, поэтому хранение нескольких экземпляров одной и той же информации будет являться избыточным.

Изображения являются неотъемлемой частью веб-сёрфинга. Вводя запрос в поисковике, пользователь получает выдачу картинок, соответствующих введённому запросу. Если некоторые изображения будут совпадать, пользователь получит меньше полезной информации, потребуется больше места на сервере для хранения изображений, больше времени на формирование поисковой выдачи.

Аналогичная проблема существует и на компьютерах пользователей. Дубликаты изображений занимают память, могут замедлять время выполнения некоторых программ.

Для решения данной проблемы необходимо отбрасывать совпадающие или похожие изображения. Данная задача является вычислительно затратной по времени, поэтому алгоритм должен выполнять сравнение за приемлемое для пользователя время работы.

Анализ предметной области

Машины создавались для того, чтобы работать с числами. Необходимость наделить их зрением возникла относительно недавно. Распознавание номерных знаков автомобилей, чтение штрих-кодов на товарах в супермаркете, анализ записей с камер наблюдения, поиск лиц на фото, создание роботов, умеющих находить и обходить

препятствия, поиск фотографий и видео, — всё это задачи, которые требуют от компьютера способности «видеть» и интерпретировать увиденное. Набор методов, позволяющих обучить машину извлекать информацию из изображения — будь то картинка или видеозапись, — называется компьютерным зрением [1]. Задача выделения значимых мест на изображении, для сравнения их с другими изображениями, является важной задачей компьютерного зрения.

Множество изображений хранится в электронном формате. При большом количестве одинаковых изображений, ожидаемый результат поиска необходимого изображения заметно ухудшается, а именно: время формирования поисковой выдачи увеличивается, количество уникальной информации уменьшается. Если рассматривать поисковики, то картинки, выведенные на запрос «*Ipad*», могут быть почти одинаковыми (рис. 1).

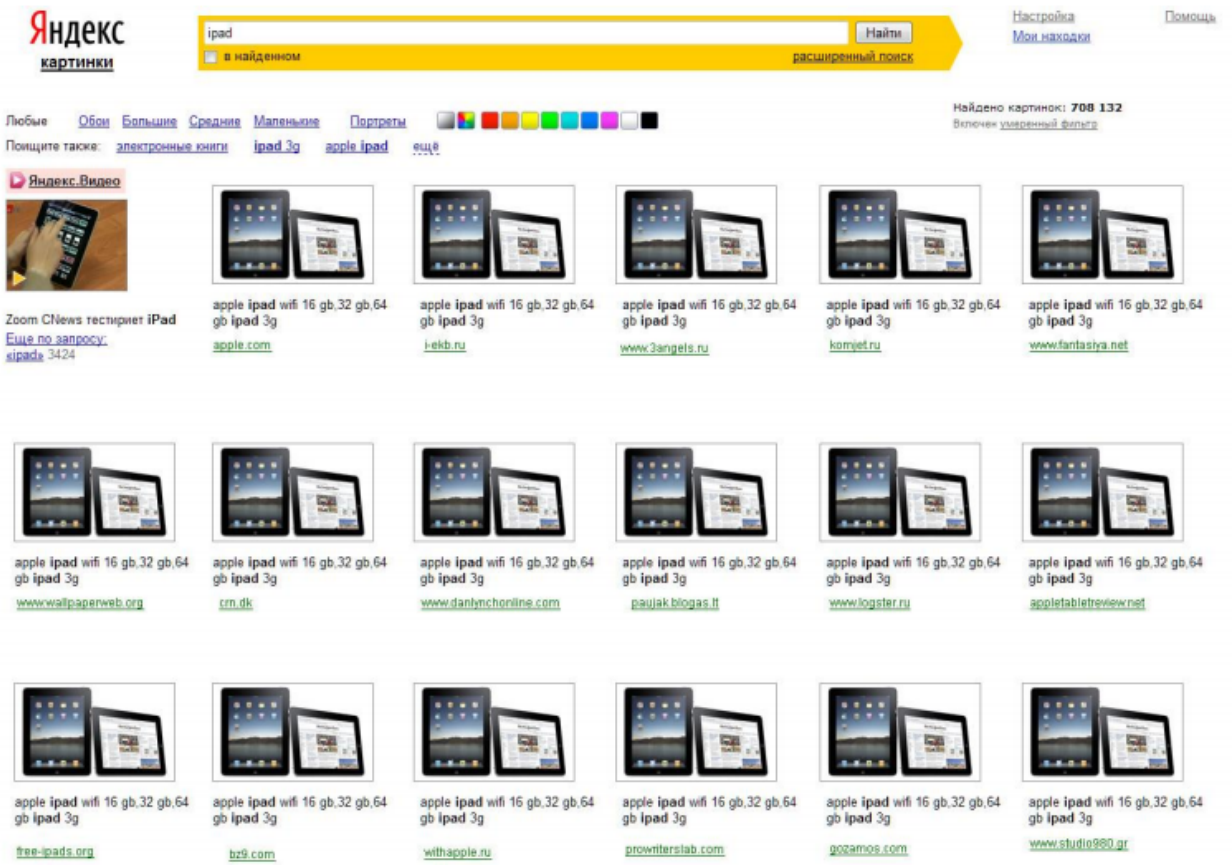


Рис. 1. Поисковая выдача без кластеризации дубликатов [2]

Данный результат не является информативным, 17 из 18 изображений оказались одинаковыми, при этом каждая картинка в отдельности действительно является правильным результатом выдачи на запрос.

Для улучшения качества поиска необходимо производить выдачу с кластеризацией дубликатов (рис. 2). Каждая картинка также является правильным результатом выдачи, но теперь она также является уникальной картинкой в коллекции. Кластеризация дубликатов не только позволяет сделать запрос более информативным, она также позволяет уменьшить размер базы изображений и ускорить вычисления на сервере – чем меньше изображений в базе, тем меньше необходимо произвести поисков по мета описанию.

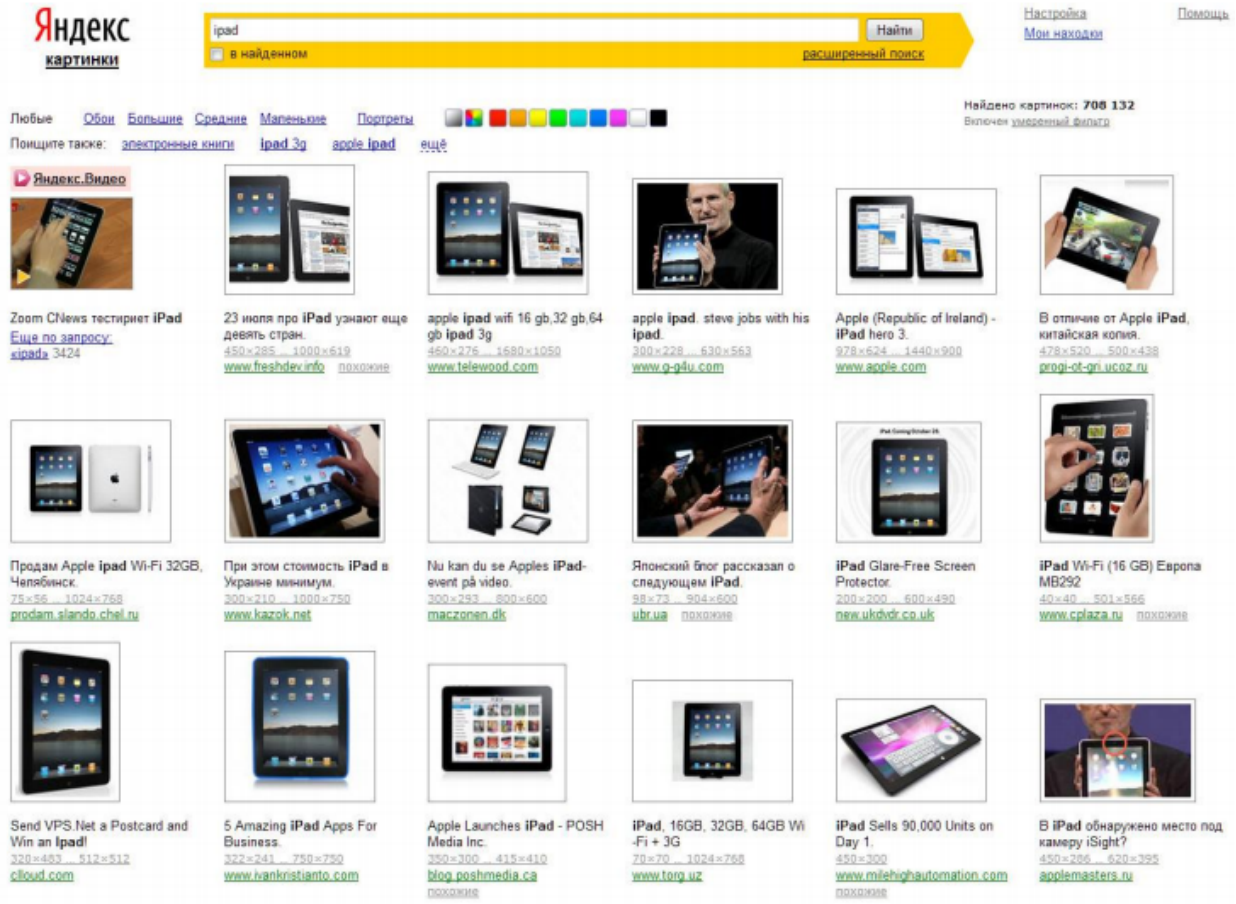


Рис. 2. Поисковая выдача с кластеризацией дубликатов [2]

Данная проблема присуща не только поисковым серверам. На пользовательских компьютерах также может находиться большое количество дубликатов изображений. Существует несколько способов провести кластеризацию, определить сходство изображений. У каждого способа свои плюсы и минусы, идеального решения проблемы не существует.

Данная работа посвящена разработке метода выделения похожих изображений на основе применения *SIFT* дескрипторов.

Дескриптор *SIFT*

Дескриптор *SIFT* является самым используемым дескриптором. На его основе построено множество алгоритмов. Данный дескриптор выполняет поиск ключевых точек и строит вектора признаков.

Для формирования дескриптора *SIFT* сначала вычисляются значения магнитуды и ориентации градиента в каждом пикселе, принадлежащем окрестности особой точки размером 16x16 пикселей. Магнитуды градиентов при этом учитываются с весами, пропорциональными значению функции плотности нормального распределения с математическим ожиданием в рассматриваемой особой точке и стандартным отклонением, равным половине ширины окрестности. В каждом квадрате размером 4x4 пикселя вычисляется гистограмма ориентированных градиентов путем добавления взвешенного значения магнитуды градиента к одному из 8 бинов гистограммы. Чтобы уменьшить различные "граничные" эффекты, связанные с отнесением похожих градиентов к разным квадратам используется билинейная интерполяция: значение магнитуды каждого градиента добавляется не только в гистограмму, соответствующую квадрату, к которому данный пиксель относится, но и к гистограммам, соответствующим соседним квадратам. При этом значение магнитуды добавляется с весом, пропорциональным расстоянию от пикселя, в котором вычислен данный градиент, до центра соответствующего квадрата. Все вычисленные гистограммы объединяются в один вектор размерностью 128 компонент.

Данный дескриптор инвариантен относительно поворота, масштабирования, сдвигов, частично инвариантен изменению угла обзора и яркости. Перечисленные преимущества позволяют использовать полученные дескрипторы для сравнения изображений на схожесть.

Описание метода выделения схожести изображений

В разработанном методе для поиска ключевых точек используется *DoG*-детектор, а для описания ключевых точек *SIFT*-дескриптор.

Алгоритм сравнения изображений:

1. Выделить особые точки изображения.
2. Определить направление градиента каждой выделенной точки.
3. Построить дескриптор каждой особой точки.
4. Повторить шаги 1-3 для второго изображения.
5. Сравнить все вычисленные дескрипторы двух изображений между собой.

Выделение особых точек

Алгоритм *SIFT* [3] использует детектор *DoG* [4] для поиска локальных точек. Данный детектор строит пирамиды гауссианов и разностей гауссианов. Гауссианом является изображение, размытое гауссовым фильтром:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y),$$

где L – значение гауссиана в точке с координатами (x, y), $*$ - операция свёртки, I – значение пикселя с координатами (x, y), G – гауссово ядро для размытия изображения:

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

где σ - величина размытия.

Пирамида гауссианов состоит из нескольких уровней. Уровень пирамиды принято называть октавой. Каждая октава содержит несколько интервалов, состоящих из изображений, размытых гауссовым фильтром с разным радиусом размытия. Каждое следующее изображение октавы размыто с большим радиусом. При переходе от одной октавы к другой размеры первоначального изображения октавы уменьшаются вдвое, благодаря чему достигается инвариантность масштабированию.

Пирамида разности гауссианов также состоит из нескольких октав. Каждая октава состоит из интервалов, содержащих разность гауссианов текущего и следующего интервалов пирамиды гауссианов рассматриваемой октавы:

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y)$$
$$= L(x, y, k\sigma) - L(x, y, \sigma),$$

где k – шаг величины размытия, увеличивающий степень размытия изображения на каждом последующем изображении.

На каждой октаве пирамид строится задаваемое число N интервалов, дополнительно к пирамиде гауссианов для каждой октавы достраивается два интервала, для возможности получить разность интервалов для N-2 и N-1 интервалов пирамиды разности гауссианов.

Таким образом число интервалов в пирамиде разности гауссианов на единицу меньше числа интервалов пирамиды гауссианов (рис. 3).

В итоге получается гауссово масштабируемое пространство, инвариантное относительно масштабирования и не смещающее локальные экстремумы.

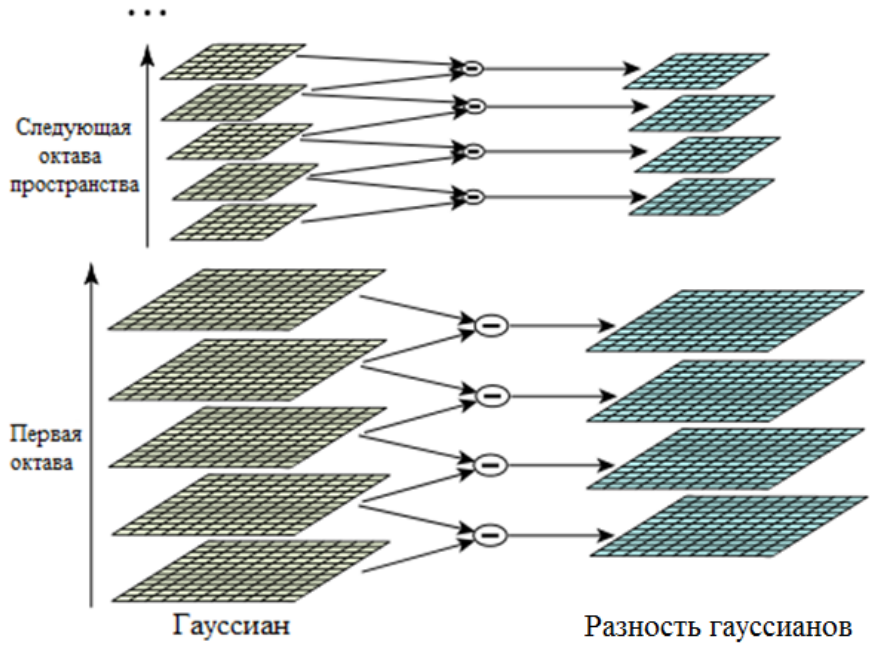


Рис. 3. Пирамиды гауссианов и разности гауссианов [3]

Масштабируемым пространством изображения является набор всевозможных, сглаженных некоторым фильтром, версий исходного изображения. Гауссово масштабируемое пространство является линейным, инвариантным относительно сдвигов и масштабирования.

После построения пирамид гауссианов необходимо выделить особые точки, которые будут являться локальными экстремумами октав пирамиды разности гауссианов (рис. 4).

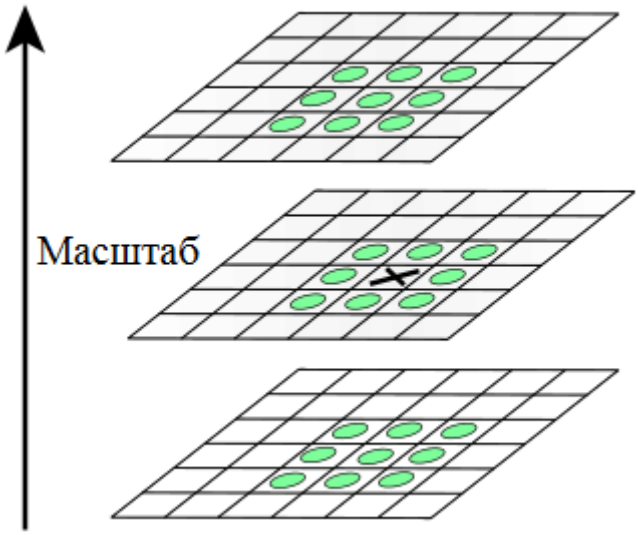


Рис. 4. Поиск локальных экстремумов в пирамиде разности гауссианов [3]

В каждом изображении каждой октавы пирамиды *DoG* ищутся точки локального экстремума. Каждая точка сравнивается с 26 соседями в текущей октаве: с восьмью в данном интервале и с девятью точками в нижнем и верхнем интервале соответственно. Точка считается точкой локального экстремума, если она больше или меньше всех соседей.

Вычисление ориентации ключевой точки

Для каждой найденной ключевой точки вычисляется её ориентация. В последующих шагах относительно этой ориентации будет проводиться построение дескриптора. Это ключевой шаг в достижении инвариантности повороту, так как в случае вращения изображения направления градиентов близлежащих точек относительно ключевой не изменятся.

Точка может иметь несколько направлений. Направление точки вычисляется в октаве с масштабом наиболее близким к масштабу ключевой точки. Для каждой точки каждого изображения в данной октаве вычисляется ориентация:

$$\theta(x,y) = tan^{-1}\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}$$

После вычислений строится гистограмма ориентаций, которая формируется из ориентаций близлежащих точек. Гистограмма имеет 36 делений, покрывающих промежуток в 360 градусов. Ориентация ключевой точки лежит в промежутке, покрываемом максимальной величиной гистограммы.

Построение дескрипторов

Предыдущие операции назначили положение, масштаб и ориентацию каждой ключевой точке. Эти точки являются повторяемыми в локальной области и обеспечивают инвариантность перечисленным параметрам. Следующим шагом является вычисление дескриптора для окрестности ключевой точки, благодаря чему к инвариантным характеристикам точки добавляется частичная инвариантность освещению и точке обзора. Дескриптор вычисляется на гауссиане, ближайшем по масштабу к ключевой точке. Этот дескриптор представляет собой локальную гистограмму направлений градиентов изображения. Окрестность характерной точки делится на четыре квадратных сектора. В каждом пикселе внутри каждого сектора вычисляется градиент изображения, его направление и модуль. Затем модули градиентов умножаются на вес, экспоненциально убывающий с удалением от точки интереса. Смысл применения веса заключается в том,

чтобы избежать резких изменений значения дескриптора при небольших изменениях положения окна, а также в том, чтобы градиенты, удаленные от центра дескриптора, вносили меньший вклад в его значение, поскольку градиенты на периферии окрестности точки интереса наименее устойчивы при геометрических преобразованиях изображения. По каждому сектору собирается гистограмма направлений градиентов, причем каждое вхождение взвешивается модулем градиента. Дескриптор *SIFT* представляет собой вектор, полученный из значений всех элементов гистограмм направлений, и состоит из 128 компонент. Дескриптор нормируется, чтобы повысить его устойчивость к изменениям яркости [5]. Таким образом для изображения будет получен набор дескрипторов.

Сравнение изображений

Чтобы сравнить два изображения, необходимо узнать количество совпадающих ключевых точек. Для этого необходимо найти эвклидово расстояние между каждый точкой одного изображения и каждой точкой другого изображения. Пара точек, у которых эвклидово расстояние минимально, является парой совпадающих точек. Задав некоторой порог минимального значения, можно получить процент совпадающих точек, исходя из которого можно узнать о степени схожести двух изображений.

Исследование алгоритма

Разработанный программный комплекс требователен к системным ресурсам и быстродействию системы в целом. Это связано с вычислительной сложностью алгоритма выделения особых точек и построения их дескрипторов. Время работы алгоритма может зависеть от размеров сравниваемых изображений. На качество сравнения двух изображений могут повлиять различные преобразования, такие как сдвиг, поворот, масштабирование, изменение яркости, перекрытие и обрезка изображения. Необходимо сравнить изображения и проанализировать полученные результаты. В качестве изображений для исследования зависимости качества и времени сравнения от преобразований было выбрано изображение шахматных фигур. Для каждого преобразования был создан дубликат исходного изображения, и к нему применено исследуемое преобразование.

Растровые изображения невозможно хранить в электронном формате не в прямоугольной форме. Единственным вариантом использования таких изображений является их увеличение, с вытекающим добавлением некоторой области. Разные редакторы могут добавлять область разных цветов и даже с некоторой штриховкой.

Поэтому изображение, повернутое на угол некратный 90 градусам, не будет являться дубликатом исходного.

Целью исследования является: определение влияния различных преобразований на качество и время выполнения сравнения двух изображений. Исследование проводилось на выборке в 80 изображений.

Преобразование	Интервал степени схожести изображений	Зависимость времени выполнения сравнения
Масштабирование	50-94 %	Зависит
Поворот	29-100 %	Не зависит
Перекрытие	62-91 %	Не зависит
Обрезка	48-89 %	Время уменьшается
Сдвиг	50-94 %	Время увеличивается
Линейное изменение яркости	60-87 %	Не зависит
Нелинейное изменение яркости	30-67 %	Не зависит

Результаты исследования

По результатам исследования можно сделать следующие выводы:

1. Разработанный метод хорошо работает для определения схожести изображений, повернутых на угол кратный 90 градусам. Метод не инвариантен относительно поворота отличного от 90 градусов.
2. Сравнение с изображением с масштабом, измененным более чем в 2-2,5 раза и потерей качества, показывает низкую меру схожести. В остальных случаях метод инвариантен относительно масштабирования.
3. Метод инвариантен при небольших перекрытиях изображения. При перекрытии более чем половины изображения инвариантность не достигается.
4. При обрезке изображения разработанный метод показывает большую степень схожести. Однако при сравнении с вырезанной областью, меньшей половины изображения, степень схожести определяется как малая.
5. Метод инвариантен относительно сдвига. Однако при заполнении добавляемых областей инвариантность не достигается. Это связано с появлением дополнительных ключевых точек в добавленных областях.

6. Линейное изменение яркости уменьшает процент схожести изображений, однако при изменении яркости в пределах +-40 % изображения определяются как схожие.
7. Разработанный алгоритм не инвариантен нелинейному изменению яркости. Это связано с изменением направлений градиентов окрестностей ключевых точек и последующими изменениями в построении дескриптора.
8. Время работы алгоритма зависит от размеров изображений. Следовательно, время работы будет зависеть от следующих преобразований: масштабирование, обрезка изображения, сдвиг и поворот на угол отличный от 90 градусов.

Заключение

В ходе работы были выполнены следующие задачи:

- Проведен анализ существующих способов сравнения изображений
- Разработан метод сравнения изображений, инвариантный повороту, масштабированию, сдвигу, перекрытию и обрезке изображения
- Проведено исследование разработанного метода на основе различных преобразований изображения
- Разработан программный продукт, позволяющий сравнивать изображения;

Разработанный метод выделения схожести изображений на основе SIFT дескрипторов инвариантен или частично инвариантен самым используемым и популярным преобразованиям, что является значимым преимуществом по сравнению с другими методами сравнения.

В качестве дальнейшего развития можно выделить следующие направления:

1. Улучшить определение схожести при нелинейном изменении яркости, масштабировании большим или меньшим чем в 2.5 раза.
2. Ускорить время сравнения изображений.

Список литературы

1. Компьютерное зрение. Блог Яндекса. Режим доступа: http://blog.yandex.ru/post/80564/ (дата обращения 11.10.14).
2. Яндекс технологии. Кластеризация дубликатов в Яндексе. Режим доступа: http://tech.yandex.ru/events/yasubbotnik/ekb-jun-2011/talks/226/ (дата обращения 11.10.14).
3. David G. Lowe. Computer Distinctive Image Features from Scale-Invariant Keypoints. Available at: http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf, accessed 11.10.14.

4. Microsoft Research. Локальные особенности изображений. Режим доступа: http://courses.graphicon.ru/files/courses/vision/2011/lectures/cv2011_05_features.pdf (дата обращения 11.10.14).
5. Международный научный центр. Поиск похожих фотографий в базе данных. Режим доступа: http://mm-dsp.com/ (дата обращения 11.10.14).