

06, июнь 2016

УДК 004.932

Использование нейросети и SIFT дескрипторов для сравнения изображений

Савонин А.И., студент

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

Научный руководитель: Rogozin O.V., к.т.н, доцент

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

bauman@bmstu.ru

Введение

В данной статье представлена система для сравнения изображений – данная задача является актуальной не только для использования в веб-сервисах, но и для домашних ПК.

Количество информации стремительно растет. Часть информации представлена изображениями. Но далеко не вся информация является уникальной, поэтому хранение нескольких экземпляров одной и той же информации будет являться избыточным.

Хранение и выдача неуникальной информации снизит качество поисковой выдачи, замедлит работу алгоритмов, затребует большее количество памяти на серверах. Для решения данной проблемы необходимо отбрасывать совпадающие или похожие изображения. Стандартным подходом для сравнения изображений является подход, при котором сравниваются дескрипторы данных изображений.

Эта задача является вычислительно затратной по времени, поэтому алгоритм должен выполнять сравнение за приемлемое для пользователя время работы. Применение нейронных сетей может позволить улучшить скорость и качество сравнения изображений.

Существуют различные подходы, позволяющие решить задачу сопоставления изображений. Каждый подход обладает своими плюсами и минусами. Нет универсального метода, решающего все задачи компьютерного зрения в полном объеме

Данная работа посвящена разработке метода сравнения изображений с использованием SIFT дескрипторов и нейронных сетей.

Метод сравнения изображений на основе SIFT дескрипторов

В предложенном методе для поиска ключевых точек используется *DoG*-детектор [2], а для описания ключевых точек *SIFT*-дескриптор.

Алгоритм сравнения изображений:

1. Выделить особые точки изображения.
2. Определить доминирующее направление градиента особой точки.
3. Построить дескриптор каждой особой точки.
4. Шаги 1-3 выполнить для обоих сравниваемых изображений.
5. Сравнить все вычисленные дескрипторы двух изображений между собой.

Поиск ключевых точек

Для выделения особых точек используется *DoG* [2] детектор. По найденным точкам строится *SIFT* [1] дескриптор. *DoG* 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 слоёв. Слой называют уровнем или 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:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma), \end{aligned}$$

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.

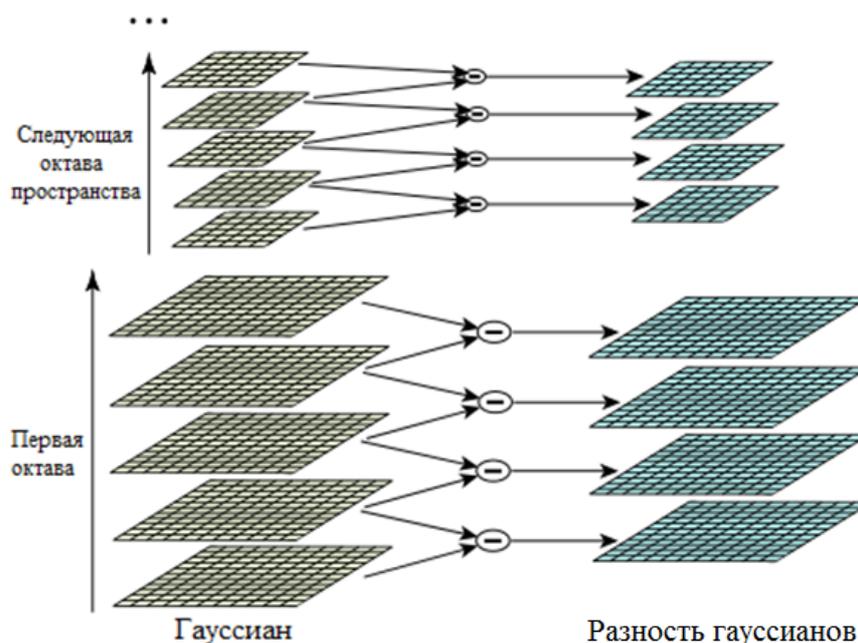


Рис. 3. Пирамиды гауссианов и разности гауссианов [1]

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).

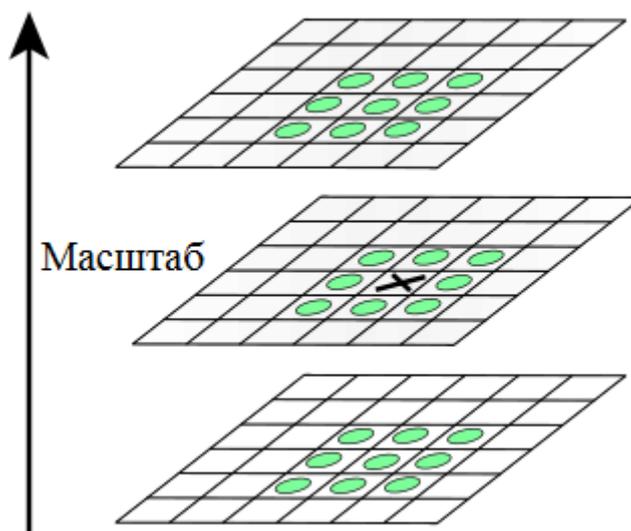


Рис. 4. Поиск локальных экстремумов в пирамиде разности гауссианов [1]

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. Такие точки называются особыми точками изображения.

Построение *SIFT* дескриптора

Дескриптор *SIFT* является самым популярным дескриптором изображений. Основное его отличие от других дескрипторов, это инвариантность самым частым преобразованиям изображений.

Для построения *SIFT* дескриптора вначале необходимо вычислить значения магнитуды и ориентации градиента в каждом из пикселей, принадлежащих окрестности ключевой 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.

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 [3]. 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 каждой точкой одного изображения и каждой точкой другого изображения [4]. 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.

Использование нейросети

Описанный метод является одним из самых используемых методов для сравнения изображений. Его используют такие компании как Яндекс и Google. Однако в данной статье предлагается усовершенствованный метод, основанный на SIFT дескрипторах с использованием нейронной сети.

В классическом алгоритма размер окна, в котором вычисляются градиенты и магнитуды, равен 16x16 пикселей. В разрабатываемом методе для нейронной сети будет использоваться окно 18x18 пикселей. Целевая функция нейросети – это дескриптор изображения. Размерность SIFT дескриптора равна 128, поэтому выходной слой нейронной сети состоит из 128 выходов. Будет использован один скрытый слой нейронной сети. Любая нейронная сеть состоит из входного слоя и выходного слоя. Соответственно подаются независимые и зависимые переменные. Входные данные преобразуются нейронами сети и сравниваются с выходом. Если отклонение больше заданного, то специальным образом изменяются веса связей нейронов между собой и пороговые значения, нейронов. Снова происходит процесс вычислений выходного значения и его сравнение с эталоном. Если отклонения меньше заданной погрешности, то процесс обучения прекращается. Помимо входного и выходного слоев в многослойной сети существуют так называемые скрытые слои. Они представляют собой нейроны, которые не имеют непосредственных входов исходных данных, а связаны только с выходами входного слоя и с входом выходного слоя.

Таким образом будет использован усовершенствованный SIFT дескриптор, для которого существует возможность обучения, с помощью метода обратного распространения ошибки. Улучшенный дескриптор позволит строить более точные вектора признаков, что повысит качество сравнения, путем увеличения процента схожести сравниваемых изображений.

Исследование алгоритма

Разработанный программный комплекс требователен к системным ресурсам и быстрдействию 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. Также был проведен анализ работы обычного метода сравнения изображений, использующего классические SIFT дескрипторы, и метода с использованием нейросети. Анализ показал, что использование улучшенного метода позволяет увеличить процент схожести, при наложении шумов или большей разности в яркости.

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.

Зависимость степени схожести и времени сравнения от преобразования

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

Заключение

В ходе работы были выполнены следующие задачи:

- Проведен анализ существующих способов сравнения изображений
- Разработан метод сравнения изображений с использованием SIFT дескрипторов и нейросети
- Проведено исследование разработанного метода на основе различных преобразований изображения
- Разработан программный продукт, позволяющий сравнивать изображения;

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

Список литературы

- [1]. David G. Lowe. Computer Distinctive Image Features from Scale-Invariant Keypoints. Available at: <http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf>, accessed 18.02.16.
- [2]. Microsoft Research. Локальные особенности изображений. Режим доступа: http://courses.graphicon.ru/files/courses/vision/2011/lectures/cv2011_05_features.pdf, (дата обращения 18.02.16).
- [3]. Международный научный центр. Поиск похожих фотографий в базе данных. Режим доступа: <http://mm-dsp.com/> (дата обращения 18.02.16).
- [4]. David G. Lowe. Object recognition from local scale-invariant features // International Conference on Computer Vision (Corfu, Greece, September 1999). P. 1150-1157.