

АЛГОРИТМ СОПОСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ ПО КЛЮЧЕВЫМ ТОЧКАМ ПРИ МАСШТАБИРУЕМОСТИ И ВРАЩЕНИИ ОБЪЕКТОВ

Д.Е. Супрун

Dianasuprun91@gmail.com

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Рассмотрен способ сопоставления изображений по ключевым точкам путем извлечения отличительных, инвариантных относительно масштаба и вращения характеристик. Описан алгоритм извлечения характеристик для эффективного сопоставления различных позиций объекта при значительном диапазоне искажения, изменении в 3D-поле зрения, наличии шума и при изменении освещенности. Приведен метод масштабно-инвариантного преобразования характеристик (SIFT) для сопоставления изображений. Предложены способы реализации метода на примере построения пирамиды гауссианов и пирамиды разностей гауссианов. Рассмотрен алгоритм нахождения точки локального экстремума, выделения ключевых точек, построения вектора признаков и сопоставлены локальные дескрипторы для дальнейшего сравнения пары изображений в условиях вращения объектов, перекрытий, изменения масштаба, точки съемки и освещенности

Ключевые слова

Сопоставление изображений, инвариантные характеристики, пирамида гауссианов, пирамида разности гауссианов, ключевая точка, локальный экстремум, вектор признаков, дескриптор ключевой точки

Поступила в редакцию 28.04.2016

© МГТУ им. Н.Э. Баумана, 2016

Задача сопоставления изображений возникает тогда, когда необходимо создать панораму, реконструировать трехмерную модель объекта, распознать объекты и т. д. Существуют методы сопоставления изображений, основанные на сравнении знаний об изображениях. Идея таких методов заключается в том, что для каждой точки изображения вычисляют значения определенной функции, на основании которых можно приписать изображению требуемую характеристику. Тогда задача сравнения изображений сводится к задаче сравнения полученных характеристик. Достоинства методов — простота использования, недостаток методов — они работают только в идеальных ситуациях; в случаях, когда на изображениях появляются шумы или изменения масштаба, применение этих методов становится затруднительным, поскольку в характеристику изображения вносит вклад каждая точка [1]. Для устранения этого недостатка необходимо из всего множества точек выделять особые «ключевые» точки для их дальнейшего сравнения. На этом принципе основан *метод сопоставления изображений по ключевым точкам* (рис. 1).

Идея метода заключается в том, что изображение заменяют некоторой моделью, состоящей из его ключевых точек. Введем несколько понятий, связанных с разработкой и применением этого метода.



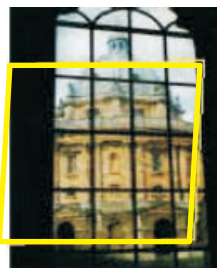
a



б



в



г

Рис. 1. Случаи применения метода сопоставления изображений по ключевым точкам:

a — изменение масштаба; *б* — изменение точки съемки; *в* — изменение освещения;
г — перекрытие

Особой, или ключевой точкой называют такую точку изображения, которая с большей долей вероятности будет находиться на другом изображении того же объекта.

Детектор — метод извлечения особых точек из изображения, который должен обеспечивать инвариантность нахождения одних и тех же ключевых точек относительно преобразований изображений. Детектор позволяет определить координаты ключевых точек. Чтобы установить, какая ключевая точка одного изображения соответствует ключевой точке другого изображения, необходимо использовать дескриптор. *Дескриптор* — идентификатор ключевой точки, который выделяет ее из множества остальных ключевых точек. Дескрипторы должны обеспечивать инвариантность нахождения соответствия между ключевыми точками относительно преобразований изображений [3].

Следует отметить, что ключевые точки инвариантны относительно изменения масштаба изображения и вращения, а также частично инвариантны к изменению освещенности и 3D-точки зрения камеры. Такие точки применимы как в пространственных, так и в частотных областях, снижая вероятность сбоев при преобразованиях, помехах и шуме. Большое число характеристик может быть извлечено из изображений с помощью эффективных алгоритмов. Кроме того, характеристики имеют явные отличительные особенности, что позволяет с высокой степенью вероятности вычислить отдельно взятую характеристику из большой базы данных характеристик для дальнейшего распознавания объектов или сцен.

Далее рассмотрим основные этапы алгоритма, используемого для получения характеристик изображения (рис. 2).

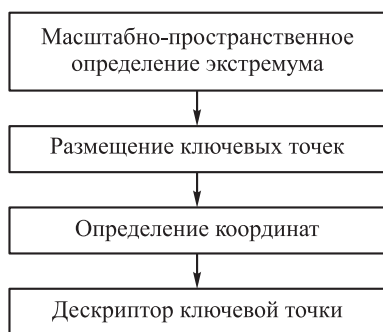


Рис. 2. Основные этапы алгоритма, используемого для получения характеристик изображения

На первом этапе расчетов выполняют анализ масштаба изображения и позиции. Эффективность метода достигается использованием функции разности гауссианов для определения потенциально интересных точек, инвариантных относительно масштаба и ориентации. На втором этапе размещения ключевых точек на каждом потенциальном месте описывают детальную модель для нахождения координат и масштаба. Ключевые точки выбирают на основе их степени стабильности. На третьем этапе одну или более координат присваивают каждой ключевой точке на основе локального направления градиента изображения.

Все дальнейшие этапы выполняют над данными изображения, которые были трансформированы в соответствии с заданными координатами, масштабом и местоположением каждой характеристики, тем самым обеспечивая инвариантность относительно этих преобразований. На четвертом этапе измеряют локальные гра-

денты изображения при выбранном масштабе в области каждой ключевой точки. Их преобразуют в представление, которое позволяет на важных уровнях исказить локальную форму и изменять освещенность. Этот подход называют масштабно-инвариантным преобразованием характеристик (SIFT), так как он преобразует данные изображения в масштабно-инвариантные координаты в соответствии с локальными характеристиками. Важный аспект этого подхода состоит в том, что он генерирует большое число функций, которые плотно покрывают изображение по всему диапазону масштабов и координат. Обычное изображение размером 500×500 пикселей включает в себя около 2000 устойчивых характеристик, хотя это число зависит как от контента изображения, так и от выбора различных параметров. Число характеристик, в частности, важно для распознавания объектов, когда возможность определения маленьких объектов из беспорядочных фонов требует, чтобы с каждого объекта были правильно подобраны по крайней мере три функции для надежной идентификации [6].

Для сравнения и распознавания изображений сначала загружают SIFT-характеристики из набора соответствующих изображений и сохраняют их в базе данных. Каждую характеристику нового изображения сравнивают с характеристиками изображения из предыдущей базы данных, и находят общие сравнительные характеристики на основе евклидова расстояния вектора характеристик. Дескрипторы ключевых точек имеют сильные отличительные особенности, которые позволяют любой характеристике с высокой долей вероятности найти соответствие в большой базе данных. Однако на изображении с шумами большинство фоновых характеристик не найдут соответствие в базе данных, увеличивая вероятность ложных соответствий. Верные соответствия могут быть отсортированы из всего набора соответствий путем определения подмножества ключевых точек, которые согласованы с объектом и его местонахождением, масштабом и ориентацией. Каждая такая группа, состоящая из трех или более характеристик и согласованная с объектом и его позицией, является предметом детальной проверки. Устанавливают любые другие характеристики, соответствующие этой позиции, а несоответствующие отбрасывают. Выполняют детальный расчет вероятности, что данный набор характеристик определяет присутствие объекта, точность сравнения и число возможно неверных соответствий. Полагают, что сравнительные характеристики объектов, прошедших эти проверки, с высокой степенью точности являются верными.

На основе анализа метода сопоставления изображений по ключевым точкам предложен детальный алгоритм (рис. 3), на каждом этапе которого существуют проблемы, решаемые различными методами.

Для нахождения ключевых точек и их дескрипторов рассмотрим метод SIFT, в основе которого лежит многомасштабное представление изображений (рис. 4).

Основной момент при нахождении ключевых точек — построение пирамиды гауссианов (Gaussian) и пирамиды разностей гауссианов (Difference of Gaussian, DoG).

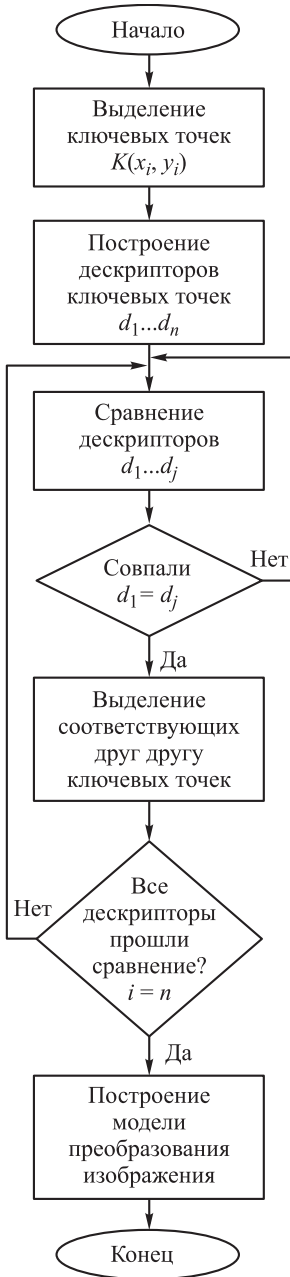


Рис. 3. Алгоритм сопоставления изображений

изображения. В связи с этим необходимо строить только некоторое их число N с определенным шагом по радиусу размытия. С тем же шагом достраивают два дополнительных гауссиана, таким образом, получают $N + 2$ гауссиана. Масштаб первого изображения следующей октавы равен масштабу изображения предыдущей октавы с номером N . Сглаживая изображение гауссовым фильтром по приведенному алгоритму, получаем пирамиду гауссианов (рис. 5).

Гауссиан — изображение, размытое гауссовым фильтром:

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

где $L(x, y, \sigma)$ — значение гауссиана в точке с координатами (x, y) ; σ — радиус размытия; $G(x, y, \sigma)$ — гауссово ядро; $I(x, y)$ — значение исходного изображения [6].

Разность гауссианов — изображение, полученное путем попиксельного вычитания одного гауссиана исходного изображения из гауссиана с другим радиусом изображения:

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

Масштабируемое пространство изображения — множество различных, сглаженных некоторым фильтром, вариантов исходного изображения. Гауссово масштабируемое пространство является линейным и инвариантным относительно сдвигов, вращений, масштаба, не смещающим локальные экстремумы, и обладает свойством полугрупп. Различная степень размытия изображения гауссовым фильтром может быть принята за исходное изображение, взятое в некотором масштабе [7].

Инвариантность относительно масштаба достигают за счет нахождения ключевых точек для исходного изображения, взятого в разных масштабах. Поэтому строят пирамиду гауссианов: все масштабируемое пространство разбивают на некоторые участки — октавы, причем часть масштабируемого пространства, занимаемого следующей октавой, в 2 раза больше части, занимаемой предыдущей. При переходе от одной октавы к другой происходит уменьшение размеров изображения в 2 раза. Каждой октаве изображения соответствует бесконечное множество гауссианов

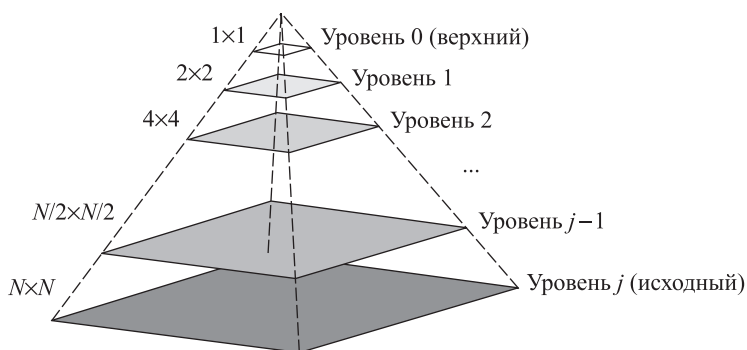
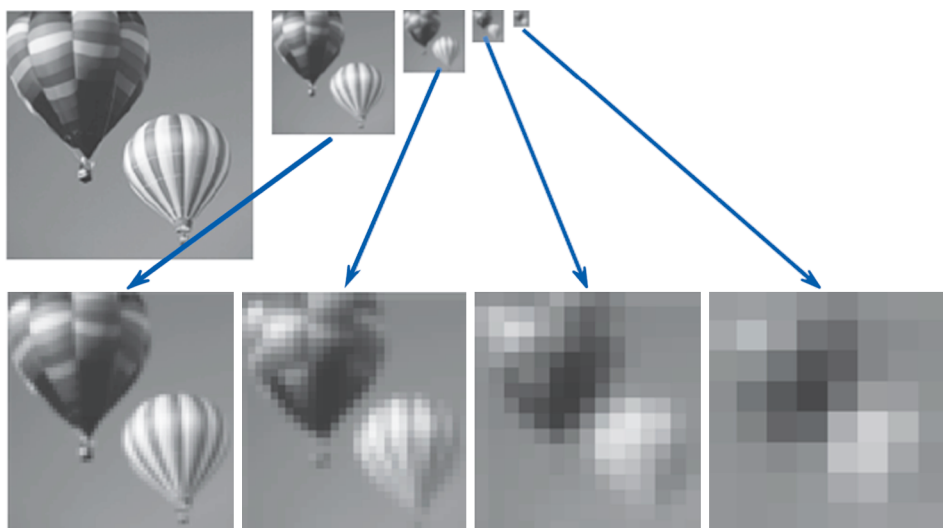
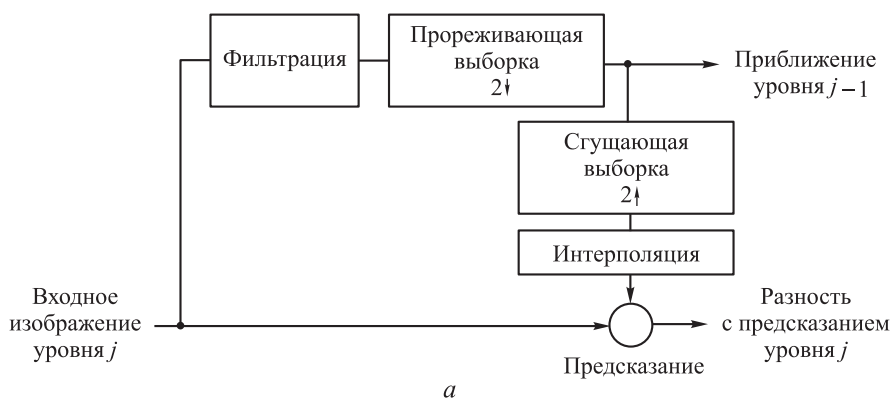


Рис. 4. Многомасштабное представление изображений



б

Рис. 5. Алгоритм построения пирамиды гауссианов (а) и результат работы алгоритма (б)

Наряду с построением пирамиды гауссианов выполняют построение пирамиды разностей гауссианов, которая состоит из разностей соседних изображений пирамиды гауссианов. Число изображений в этой пирамиде будет равно $N + 1$. Пирамиды гауссианов и разности гауссианов приведены на рис. 6. Схематично представлено, как получить пирамиду разностей гауссианов; число разностей на единицу меньше числа гауссианов и при переходе к следующей октаве размер изображений уменьшается в 2 раза.

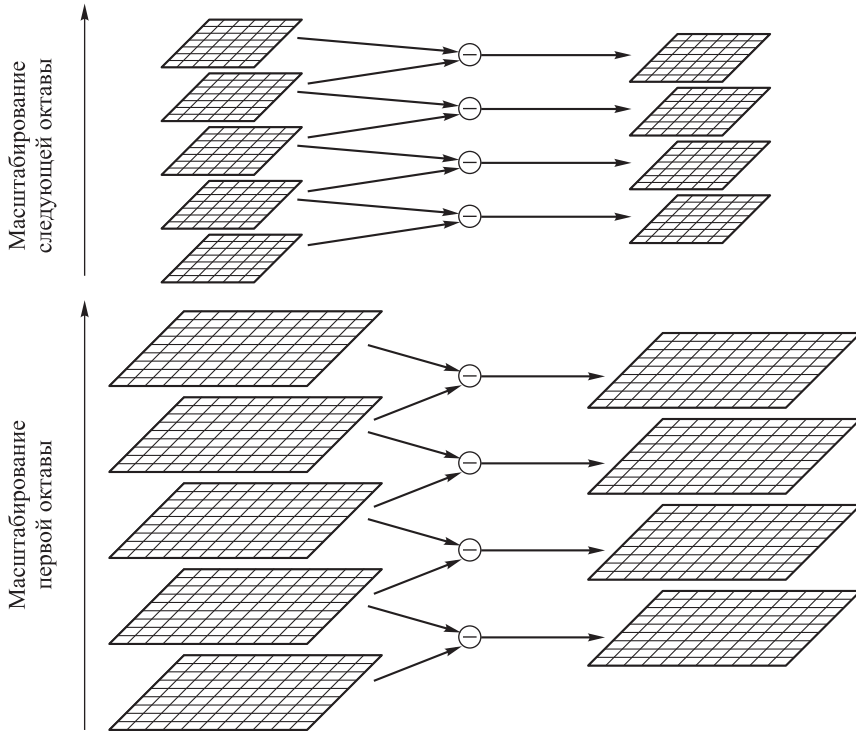


Рис. 6. Пирамиды гауссианов (а) и разности гауссианов (б)

После построения пирамид определяем: является ли точка ключевой. Точка ключевая, если она представляет собой локальный экстремум разности гауссианов.

В каждом изображении пирамиды разностей гауссианов находят точки локального экстремума. Каждую точку текущего изображения пирамиды разностей сравнивают с ее восемью соседями и с девятью соседями в пирамиде разностей, которые расположены на уровень выше и ниже в пирамиде. Если эта точка больше или меньше всех соседей, то ее принимают за точку локального экстремума, что объясняет, для чего были необходимы два дополнительных изображения в октаве. После определения точек экстремума необходимо проверить ключевые точки. Для этого следует аппроксимировать функции матрицы разности гауссианов многочленом Тейлора второго порядка, взятого в точке экстремума:

$$D(x) = D + \frac{\partial D^T}{\partial x} + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x,$$

где D — функция разности гауссианов; $x = (x, y, \sigma)$ — вектор смещения относительно точки разложения; $\frac{\partial D^T}{\partial x}$ — градиент; $\frac{\partial^2 D}{\partial x^2}$ — матрица Гессе.

Экстремум многочлена Тейлора находят вычислением производной и приравниванием ее к нулю. В результате этого получают смещение точки рассчитанного экстремума относительно точного значения функции:

$$\hat{x} = \frac{\partial^2 D^T}{\partial x^2} \frac{\partial D}{\partial x}.$$

Производные определяют по формулам конечных разностей, в итоге получают СЛАУ размерностью 3×3 относительно компонент вектора \hat{x} . Если одна из компонент вектора \hat{x} больше 0,5 (шаг сетки в этом направлении), то это означает, что точка экстремума вычислена неверно, и необходимо перейти к соседней точке в направлении указанных компонент. Для соседней точки весь алгоритм повторяют сначала. Если вышли за пределы октавы, то следует исключить данную точку из рассмотрения. Когда положение точки экстремума определено, проверяют на малость значение пирамиды разностей гауссианов в этой точке:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x}.$$

Если точка проверку не проходит, то ее исключают, так как у точки слишком маленький контраст.

На последнем этапе проверки выполняют анализ, лежит ли ключевая точка на границе какого-либо объекта или плохо освещена. Если это утверждение верно, то точку исключают из рассмотрения. Эти точки имеют большой изгиб вдоль границы и малый изгиб в перпендикулярном направлении, который определяется матрицей Гессе H размерностью 2×2 :

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}.$$

Пусть $\text{Tr}(H)$ — след матрицы, а $\text{Det}(H)$ — ее определитель, тогда

$$\text{Tr}(H) = D_{xx} + D_{yy} = \alpha + \beta;$$

$$\text{Det}(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$$

Пусть r — отношение большего изгиба к меньшему изгибу.

Точка является ключевой, если она удовлетворяет условию

$$\frac{(\text{Tr}(H))^2}{\text{Det}(H)} < \frac{(r+1)^2}{r}.$$

После определения ключевой точки необходимо установить ее ориентацию, которую находят исходя из направлений градиентов точек, соседних с ключевой точкой. Все вычисления градиентов проводят на изображении в пирамиде гауссианов с масштабом, наиболее близким к масштабу ключевой точки. Значение и направление градиента в точке с координатами (x, y) рассчитывают по формулам [3]:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2};$$

$$\theta(x, y) = \text{tg}^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right),$$

где $m(x, y)$ — градиент; $\theta(x, y)$ — направление градиента.

Для начала необходимо определить окрестность ключевой точки, в которой будут рассмотрены градиенты. Это будет окрестность, требуемая для свертки с гауссовым ядром, причем ядро будет круглым и радиус размытия для этого ядра σ составит 1,5 (масштаб ключевой точки). Для гауссова ядра действует так называемое правило трех сигм: значение гауссова ядра очень близко к нулю на расстоянии, превышающем 3σ . Следовательно, радиус окрестности равен 3σ . Направление ключевой точки находят из гистограммы направлений, которая состоит из 36 компонентов, равномерно покрывающих промежутков в 360° , и формируется следующим образом: каждая точка окрестности (x, y) вносит вклад, равный $mG(x, y, \sigma)$, в тот компонент гистограммы, который покрывает промежуток, содержащий направление градиента $\theta(x, y)$. Направление ключевой точки лежит в промежутке, покрываемом максимальным компонентом гистограммы. Значения максимального компонента (\max) и двух соседних с ним интерполируют параболой, и точка максимума этой параболы берется в качестве направления ключевой точки. Если в гистограмме есть еще компоненты, не меньше $0,8\max$, то их интерполируют аналогично и приписывают ключевой точке дополнительные направления [6].

Дескриптором может быть любой объект, но чаще всего им является некоторая информация об окрестности ключевой точки. Это связано с тем, что на маленькие области меньшее влияние оказывают эффекты искажений, некоторые изменения (изменение положения объекта на изображении, изменение сцены, перекрытие одного объекта другим, поворот) могут не повлиять на дескриптор.

В методе SIFT дескриптор — вектор, который вычисляют на гауссиане, ближайшем по масштабу к ключевой точке, и исходя из градиентов в некоторой окрестности ключевой точки. Перед определением дескриптора окрестность поворачивают на угол направления ключевой точки, за счет чего достигают инвариантность относительно поворота.

Часть изображения (рис. 7, а) представлена в виде пикселей, обозначенных квадратиками. Пиксели берутся из квадратного окна дескриптора, которое в свою очередь поделено еще на четыре равных части. Стрелочка, которая находится в

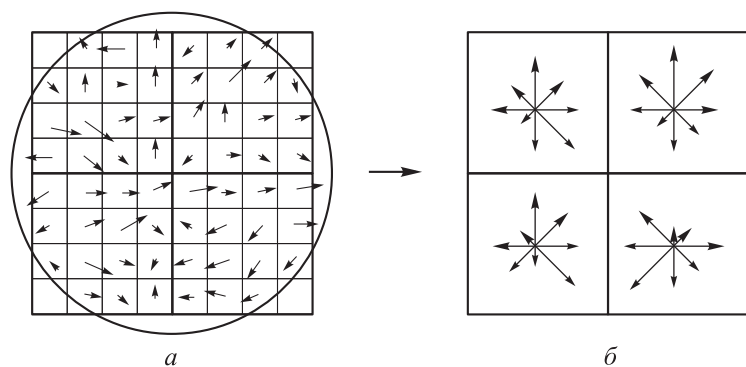


Рис. 7. Схема определения дескриптора:

a — направление градиента изображения; *б* — дескриптор ключевой точки

центре каждого пикселя, обозначает его градиент. Центр окна находится между пикселями, его необходимо выбирать как можно ближе к точным координатам ключевой точки. Круг обозначает окно свертки гауссовым ядром. Для этого ядра определено значение σ , равное половине ширины окна дескриптора. Далее значение каждой точки окна дескриптора будут умножать на значение гауссова ядра в этой точке, как на весовой коэффициент.

Дескриптор ключевой точки размерностью $2 \times 2 \times 8$ приведен на рис. 7, б. Значение 2×2 — число регионов по горизонтали и вертикали; значение 8 — число компонентов гистограммы этих регионов. Гистограммы в регионах вычисляют так же, как и гистограмму направлений с учетом того, что:

- гистограмма покрывает участок в 360° , но делит его на восемь частей, а не на 36;
- в качестве весового коэффициента берут значение гауссова ядра;
- коэффициенты трилинейной интерполяции являются весовыми коэффициентами.

Каждому градиенту в окне дескриптора можно присвоить три вещественных координаты (x, y, n) , где x — расстояние до градиента по горизонтали; y — расстояние по вертикали; n — расстояние до направления градиента в гистограмме. За точку отсчета принимают левый нижний угол окна дескриптора и начальное значение гистограммы. За единичные отрезки берут размеры регионов по горизонтали и вертикали для x и y соответственно, и градусы в компоненте гистограммы для n . Коэффициент трилинейной интерполяции определяют для каждой координаты (x, y, n) градиента как разность $l - d$, где d — расстояние от координаты градиента до середины того единичного промежутка, в который эта координата попала. Каждое вхождение градиента в гистограмму умножается на три весовых коэффициента трилинейной интерполяции [6].

Дескриптор ключевой точки состоит из всех полученных гистограмм. Размерность дескриптора (см. рис. 7) — 32 компонента ($2 \times 2 \times 8$), но на практике используют дескрипторы размерностью 128 компонентов ($4 \times 4 \times 8$). Полученный дескриптор нормализуется и готов к использованию [1].

Рассмотрим пример получения дескриптора с помощью определения массива 4×4 градиента направления из области окрестности интересующей точки 16×16 пикселей. Набор загруженных интересных точек и их областей, используемых для получения дескрипторов, представлен на рис. 8, а. Далее выбираем интересующий регион (рис. 8, б, в) и наблюдаем, как изменяется градиент до и после вращения по отношению к исходной ориентации данной области точки. Таким образом, получаем дескриптор интересующей области (рис. 8, г).

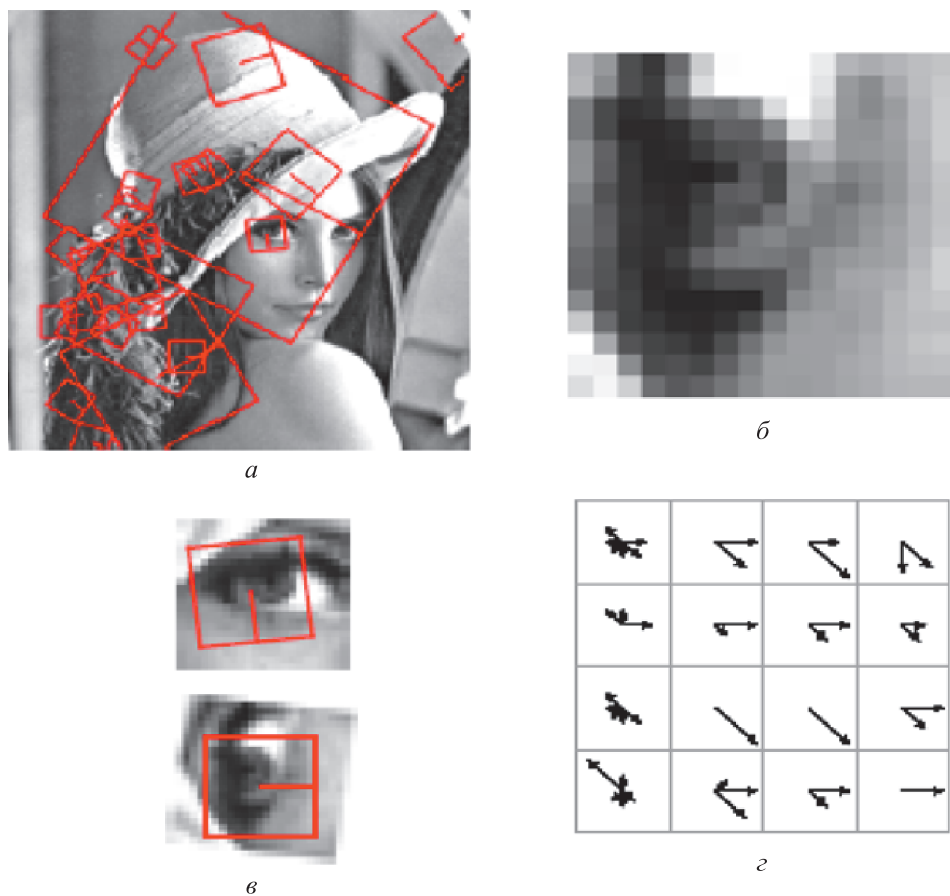


Рис. 8. Пример получения дескриптора

У дескрипторов также имеются недостатки. Не все полученные точки и их дескрипторы будут удовлетворять предъявляемым требованиям, что будет влиять на дальнейшее решение задачи сопоставления изображений. В некоторых случаях решение может быть не найдено, даже если оно существует. Несмотря на это, указанные дескрипторы хорошо работают во многих практически важных случаях.

Заключение. При сопоставлении изображений важным аспектом является инвариантность к масштабу. Изображения сопоставляют по ключевым точкам путем извлечения отличительных, инвариантных относительно масштаба и

вращения характеристик. Для этого применяют метод масштабно-инвариантного преобразования характеристик SIFT, строят векторы признаков и сопоставляют локальные дескрипторы пары изображений. Такой алгоритм предназначен для эффективного сопоставления различных позиций объекта при его вращении, значительном диапазоне искажения, изменении поля зрения, освещенности и наличии шума.

ЛИТЕРАТУРА

1. Гаганов В. Инвариантные алгоритмы сопоставления точечных особенностей на изображениях // Компьютерная графика и мультимедиа. 2009. № 7 (1). URL: http://cgm.computergraphics.ru/issues/issue17/invariant_features
2. Григорьев Ю.А., Ревунков Г.И. Банки данных. М.: Изд-во МГТУ им. Н.Э. Баумана, 2002. 318 с.
3. Супрун Д.Е., Матвеев В.А. Алгоритм создания виртуального мини-музея // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2013. № 4. С. 67–78.
4. Pope A., Lowe D. Probabilistic models of appearance for 3D object recognition // International Journal of Computer Vision. 2000. Vol. 40 (2). P. 149–167.
5. Meng Y. Implementing the scale invariant feature transform (SIFT) method // Computer Science Department University of British Columbia Vancouver, B.C., Canada, 2006.
6. Brown M., Winder S., Szeliski R. Multi-image matching using multi-scale oriented patches // Conference on Computer Vision and Pattern Recognition. 2005. P. 510–517.
7. Lowe D. Distinctive image features from scale-invariant key points // Computer Science Department University of British Columbia Vancouver, B.C., Canada, 2004.

Супрун Диана Евгеньевна — аспирантка кафедры «Системы обработки информации и управления» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Просьба ссылаться на эту статью следующим образом:

Супрун Д.Е. Алгоритм сопоставления изображений по ключевым точкам при масштабируемости и вращении объектов // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2016. № 5. С. 86–98. DOI: 10.18698/0236-3933-2016-5-86-98

IMAGE-MATCHING ALGORITHM USING KEY POINTS WITH SCALABILITY AND ROTATION OF OBJECTS

D.E. Suprun

Dianasuprun91@gmail.com

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

This article describes an image-matching algorithm using key points to extract distinctive features invariant to scale and rotation. My work presents a feature extraction algorithm for reliable comparison of the different positions of the object or scene with a significant distortion range, changes in the

Keywords

Image matching, invariant characteristics, pyramid of Gaussians, Difference of Gaussians, key point, local extremum, feature vector, key point descriptor

3D view, presence of noise and illumination changes. The scale-invariant feature transform (SIFT) method was used for matching images. The paper offers method implementation based on the pyramid of Gaussians and the Difference of Gaussian (DoG). The algorithm provides an opportunity to find the local extreme point, to detect key points, to build a feature vector and to compare local descriptors for further image pair matching under conditions of rotation, overlap, scale, change in point of shooting or lighting

REFERENCES

- [1] Gaganov V. Invariant algorithms comparison point features in images. *Komp'yuternaya grafika i mul'timedia* [Computergraphics and Multimedia], 2009, no. 7 (1) Available at: http://cgm.computergraphics.ru/issues/issue17/invariant_features
- [2] Grigor'ev Yu.A., Revunkov G.I. Banki dannyykh [Databases]. Moscow, MGTU im. N.E. Bauman Publ., 2002. 318 p.
- [3] Suprun D.E., Matveev V.A. The algorithm for creating a virtual mini-museum. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Bauman. Priborostr.* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.], 2013, no. 4, pp. 67–78 (in Russ).
- [4] Pope A., Lowe D. Probabilistic models of appearance for 3-D object recognition. *International Journal of Computer Vision*, 2000, vol. 40 (2), pp. 149–167.
- [5] Meng Y. Implementing the Scale Invariant Feature Transform (SIFT) Method. *Computer Science Department University of British Columbia Vancouver*, B.C., Canada, 2006.
- [6] Brown M., Winder S., Szeliski R. Multi-image matching using multi-scale oriented patches. *Conference on Computer Vision and Pattern Recognition*, 2005, pp. 510–517.
- [7] Lowe D. Distinctive image features from scale-invariant key points. *Computer Science Department University of British Columbia Vancouver*, B.C., Canada, 2004.

Suprun D.E. — post-graduate student of Systems of Data Processing and Control Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Suprun D.E. Image-Matching Algorithm using Key Points with Scalability and Rotation of Objects. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Bauman, Priborostr.* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.], 2016, no. 5, pp. 86–98.
DOI: 10.18698/0236-3933-2016-5-86-98