

УДК 004.93

Сравнение локальных дескрипторов при локализации контрольных точек на лице

Жиденко М.В., студент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедры «Биомедицинские технические системы»*

Научный руководитель: Самородов А.В., к.т.н., доцент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана
bmt-1@bmstu.ru*

Введение

Локальные дескрипторы представляют собой вектора признаков, построенные по отдельным фрагментам изображения. Дескрипторы нашли успешное применение во многих областях: распознавание объектов, распознавание текстуры, поиск изображений, сравнение изображений, создание панорам, распределение объектов по категориям, интеллектуальный анализ видеоданных. Практически важный интерес представляет использование локальных дескрипторов для локализации контрольных точек на изображении лица человека в условиях наличия дополнительных факторов, таких как отклонение ракурса головы от фронтального, изменение направления взгляда и пр. В данной статье приведены результаты исследования пригодности трех наиболее распространенных локальных дескрипторов для решения указанной задачи.

Локальные дескрипторы

1. Вычисление первого локального дескриптора основано на алгоритме выделения локальных текстурных признаков в окрестности КТ с помощью фильтров Гаусса [1]. Фильтр представляет собой набор из 16-ти т.н. ядер, каждое из которых чувствительно только к одному из 16-ти направлений изменения текстуры и находится как разность между двумя функциями Гаусса (рис. 1):

$$G_{pqf_k}^- = \frac{1}{2\pi\sigma^2} * \exp\left(-\frac{(p - q * \cos(f_k))^2 + (p - \sin(f_k))^2}{2\sigma^2}\right),$$

$$G_{pqf_k}^+ = \frac{1}{2\pi\sigma^2} * \exp\left(-\frac{(p + q * \cos(f_k))^2 + (p + \sin(f_k))^2}{2\sigma^2}\right),$$

где σ – СКО;

f_k – угол поворота; $f_k = k * 22,5^\circ$; $k=0\dots 15$;

$p \in [-3; 3]$, $q \in [-3; 3]$.

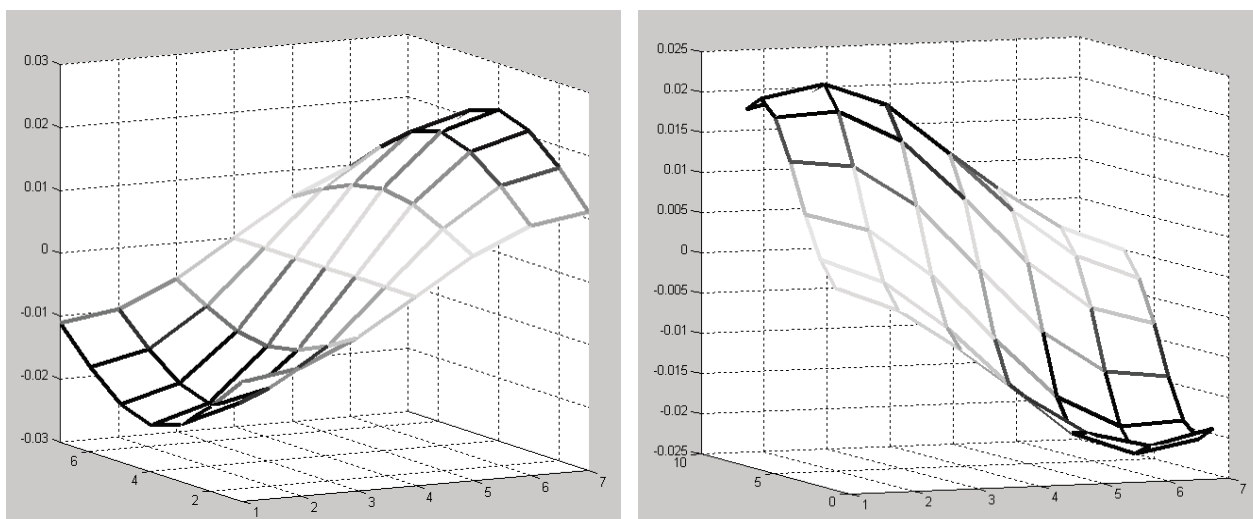


Рис. 1. Вид функции Гаусса для различных направлений: а) угол $f=0^\circ$, б) угол $f=157,5^\circ$

Далее происходит вычисление свертки по всей области интереса по следующей формуле:

$$CE_{ijfk} = \sum_{p,q} C_{i-p,j-q} * (G_{pqf_k}^- - G_{pqf_k}^+),$$

где C_{ij} представляет собой набор цветовых параметров изображения.

В оригинальной работе [1] в качестве используемых цветовых параметров изображения лица рассматриваются «тон», «насыщенность» и «яркость», однако, как показали проведенные исследования данного алгоритма, достаточно использовать однотонное изображение с 256 градациями серого. Максимальный отклик среди 16-ти значений CE_{ijfk} определяет максимальное изменение локальной границы вокруг данной точки.

Вектор текстурных признаков пикселя изображения состоит из 49 элементов. На Рисунке 2 показано, в каком виде рассматривается окрестность точки ($\Psi = 22,5^\circ$). Каждый узел сетки – это результат свертки окрестности данного узла с описанным выше фильтром. Расстояния R_0 , R_1 и R_2 выбираются исходя из практических соображений.

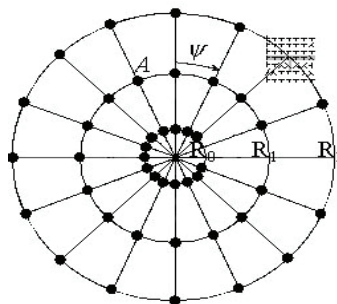


Рис. 2. Вычисление текстурных признаков контрольной точки

Сравнение вектора текстурных признаков текущей точки с шаблонным происходит путем вычисления меры K :

$$K = \sum_{i=1}^{49} th \left(\frac{1}{16} \sum_{j=1}^{16} Sg(A_{ij}^b; A_{ij}^{rw}) \right),$$

где A_{ij} является значением свертки для j -ой ориентации в конкретной области для i -ой точки входного окна;

b – шаблонный вектор текстурных признаков;

rw – текущий вектор текстурных признаков;

$$Sg(a, b) = \begin{cases} 1 \text{ if } (a > 0) \wedge (b > 0), \\ 1 \text{ if } (a = 0) \wedge (b = 0), \\ 0 \text{ в противном случае.} \end{cases}$$

$$th(x) = \begin{cases} x \text{ if } x > \text{пороговое значение,} \\ 0 \text{ в противном случае.} \end{cases}$$

Пороговое значение подобрано эмпирически и принято равным 0,2. Чем выше полученное значение меры K , тем больше схожесть окрестности рассматриваемой точки с шаблонной, полученной при обучении.

2. Следующий рассмотренный дескриптор основан на преобразовании SIFT (Scale Invariant Feature Transform) [2]. Получаемые с помощью него признаки инвариантны относительно масштаба и поворота, устойчивы к ряду аффинных преобразований, шуму, изменению в освещении. Данный алгоритм можно разделить на две части: определение «точек интереса» и построение дескрипторов окрестностей данных точек. Существует

несколько способов определения точек интереса. Алгоритм, предложенный в рамках SIFT, – один из самых известных. Он заключается в построении и использования пирамиды Гаусса изображения. Все изображения из указанной пирамиды приводятся к одному размеру, и вычисляется их разность, как показано на Рисунке 3. В качестве кандидатов точек интереса выбираются только те пиксели, которые сильно отличаются от остальных. Их поиск осуществляется путем сравнения каждого пикселя изображения с несколькими соседними на данном масштабе, с несколькими соответствующими соседями в большем и меньшем масштабе (рис. 3). Пиксель выбирается в качестве точки интереса только в том случае, если значение его интенсивности является экстремумом.

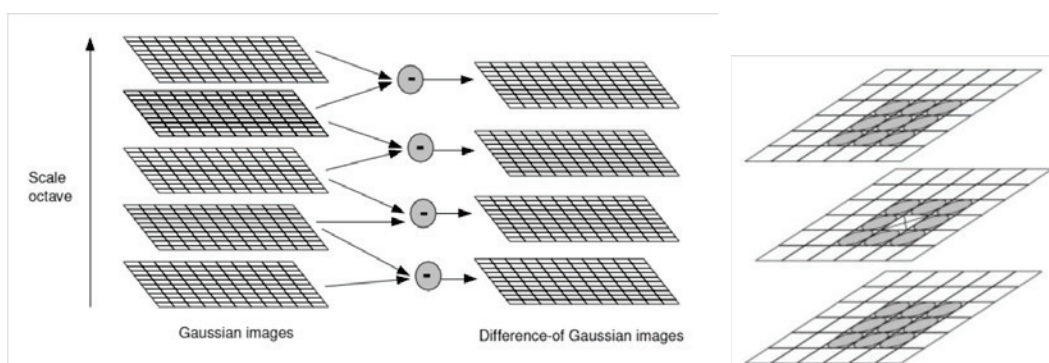


Рис. 3. Процесс нахождения точек интереса

Далее для каждой точки интереса вычисляется локальный дескриптор, характеризующий направление градиентов в пикселях некоторой его окрестности (рис. 4).

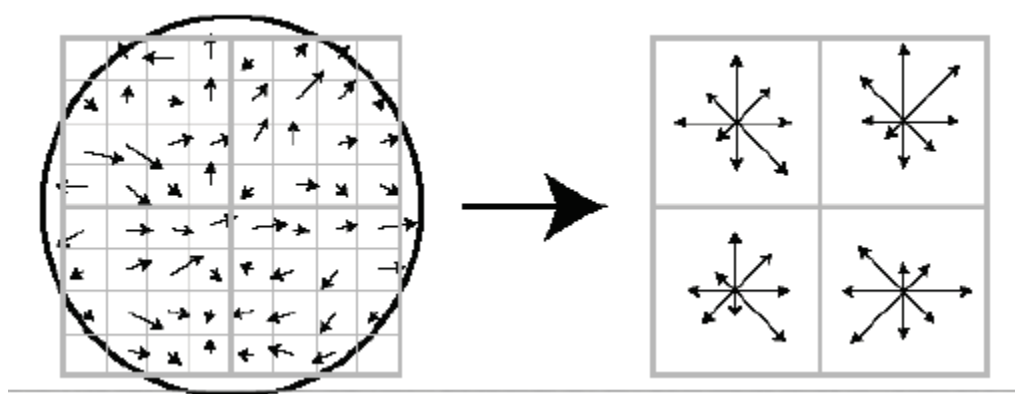


Рис. 4. Гистограммы градиентов Lowe (SIFT)

Основным недостатком SIFT дескрипторов является их высокая размерность (128 значений на точку) и большое количество точек интереса на изображении. Для

преодоления данного недостатка в работе [3] предложен метод PCA-SIFT, основанный на снижении размерности дескриптора с помощью метода главных компонент.

3. Все вышеописанные локальные дескрипторы используют общую парадигму, заключающуюся в нахождении точек интереса и построении для каждой из них дескриптора, описывающего ее окрестность. Принципиально другой подход описан в работе [4]. Он основан на свойстве повторяемости (самоподобия) фрагментов на изображении, то есть на наблюдении, что небольшие фрагменты изображения имеют свойство повторяться на нем в том же или другом масштабе. Информация о такой повторяемости в пределах некоторой области изображения формирует так называемую геометрическую разметку. С помощью данной геометрической разметки формируются самоподобные локальные дескрипторы.

Процесс вычисления данного дескриптора можно разбить на несколько этапов [4]:

- Вычисление «поверхности расстояний». В качестве «эталона» используется область размером 5x5 пикселей вокруг рассматриваемой точки. С использованием данного эталона рассчитывается «сумма квадратов расстояний между значениями интенсивностей» (SSD_q) по области размером 41x41 пикселей вокруг рассматриваемой точки.
- Вычисление «корреляционной поверхности». Выполняется нормировка поверхности расстояний SSD_q в соответствии с соотношением:

$$\overline{SSD}_q(x, y) = \frac{SSD_q(x, y)}{\max(var_{noise}, var_{auto})},$$

где var_{noise} – константа, которая указывает допустимые вариации (в цвете, освещении или допустимый шум);

$var_{auto}(q)$ – зависит от контраста фрагмента и его структуры и определяется как максимальная разница между фрагментом с центром в q и фрагментами в центрах $q_1, \dots, q_8$;

$q_1, \dots, q_8$ – соседи пикселя q .

Отсчеты результирующей «корреляционной поверхности» рассчитываются как экспоненциальные значения $\overline{SSD}_q(x, y)$:

$$S_q(x, y) = e^{-\overline{SSD}_q(x, y)}.$$

- Построение дескриптора. Полученная «корреляционная поверхность» разбивается на 20 секторов и 4 значения радиуса в логарифмических полярных координатах. В каждой области ищется максимум корреляционной поверхности, который является отсчетом дескриптора (рис. 5).

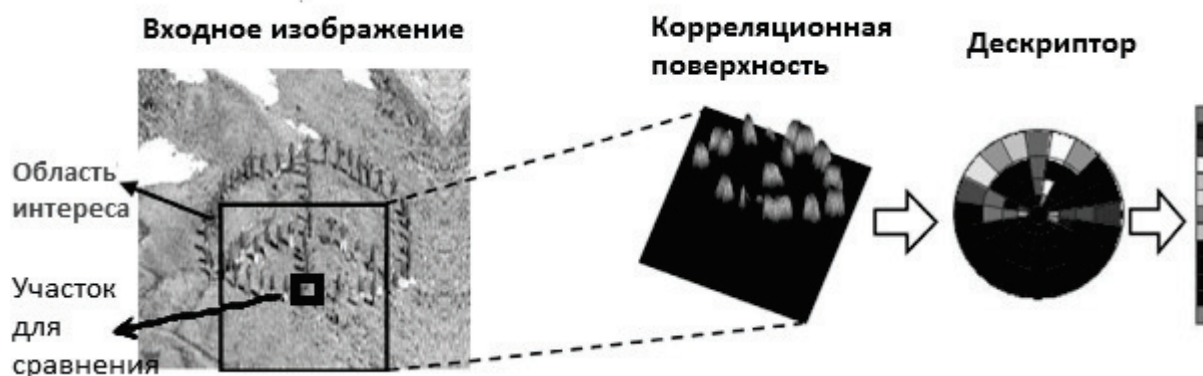


Рис. 5. Вычисление самоподобного локального дескриптора [4]

Мера подобия двух дескрипторов вычисляется по формуле, приведенной в работе [5]:

$$P(d_y^i, d_x^i) = e^{-(d_y^i - d_x^i)^T * S_D^{-1} * (d_y^i - d_x^i)},$$

где S_D – константная матрица ковариаций, которая определяет допустимое отклонение значений дескрипторов.

Тестирование дескрипторов

Тестирование проводилось на двух базах изображений. Одна из них – IMM Face Database [6] – представляет собой набор фотографий 40 различных людей, снятых под разными ракурсами и при изменении освещении. Вторая – собранная в НИИЦ БТ МГТУ им. Н.Э. Баумана – видеофрагменты 4 разных людей, совершающих различные движения глазами и головой: моргание, взгляд вверх, влево и вправо, наклоны головы вперед и назад, отклонения головы вправо и влево, повороты головы вправо и влево. В качестве контрольных точек были выбраны внешние углы глаз, вершина носа, углы рта. Оценка проводилась путем вычисления евклидова расстояния между координатами найденных точек и размеченными вручную координатами. Для базы IMM обучение проводилось на изображении того же человека, снятого фронтально без эмоций. Для второй выборки обучение проводилось на первом кадре видеофрагмента для конкретного движения. Все

алгоритмы реализованы в среде разработки Matlab. Результаты тестирования для IMM представлены в Таблицах 1–3. Числовые значения, представленные в таблицах – евклидовы расстояния между найденными положениями контрольных точек и действительными, установленными визуально.

Таблица 1.

Не нейтральное выражение лица (улыбка)

	Правый глаз	Левый глаз	Правый угол рта	Левый угол рта	Все точки вместе
Гауссовы фильтры	3,42	3,18	6,04	4,84	17,48
SIFT	3,93	3,71	8,07	8,92	24,63
Самоподобный дескриптор	4,16	3,83	5,98	6,56	20,52

Таблица 2.

Поворот головы вправо ~30-40°

	Правый глаз	Левый глаз	Правый угол рта	Левый угол рта	Все точки вместе
Гауссовы фильтры	5,58	9,54	3,8	3,24	22,16
SIFT	7,65	10,32	8,95	8,48	35,39
Самоподобный дескриптор	4,67	9,65	5,61	3,6	23,54

Таблица 3.

Изменение освещения

	Правый глаз	Левый глаз	Правый угол рта	Левый угол рта	Все точки вместе
Гауссовы фильтры	2,99	2,62	2,78	2,52	10,90
SIFT	5,87	5,23	4,27	4,18	19,54
Самоподобный дескриптор	3,77	3,82	2,99	2,62	13,19

Результаты тестирования дескрипторов на второй базе изображений лица представлены в Таблице 4.

Таблица 4.

Результаты тестирования базы изображений с различными движениями

Номер движения	SIFT	Гауссовы фильтры	Самоподобный дескриптор	Движение
1	22,3	23,6	23,2	моргание
2	18,5	20,3	21,7	взгляд вверх
3	17,3	20,5	19,2	взгляд влево
4	17,6	18,7	19,2	взгляд вправо
5	26,5	37,6	32,7	отклонение головы вправо
6	25,7	35,1	33,8	отклонение головы влево
7	43,0	48,9	50,9	поворот головы вправо
8	47,2	52,1	50,6	поворот головы влево
9	58,8	64,3	71,2	наклон головы вперед
10	42,3	48,5	42,5	наклон головы назад

Обсуждение результатов и выводы

Рассматриваемые в данной статье дескрипторы показали достаточно высокие результаты при их использовании в задаче локализации контрольных точек лица в различных условиях. При тестировании на отдельных фотографиях (база изображений IMM) дескриптор, основанный на вычислении серии сверток с гауссовыми фильтрами, показал заметно лучшие результаты; при тестировании на динамических видеофрагментах дескриптор на основе SIFT оказался более устойчивым. По скорости в Matlab самоподобный дескриптор – самый медленный, быстроедействие остальных двух сопоставимо, однако недостаточно для обработки изображений в режиме реального времени.

Список литературы

1. Gao X.W., Anishenko S., Shaposhnikov D. High-precision Detection of Facial Landmarks to Estimate Head Motions Based on Vision Models. Journal of Computer Science, Vol.3, No.7. 2007. С. 528 – 532
2. L. David, Object recognition from local scale-invariant features // Proceedings of the International Conference on Computer Vision
3. Y. Ke, R. Sukthankar, PCA-SIFT: A More Distinctive Representation for Local Image Descriptors

4. M. Irani, E. Shechtman, Matching Local Self-Similarities across Images and Videos // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007
5. O. Boiman, M. Irani, Detecting irregularities in images and in video // ICCV, Beijing, 2005
6. The IMM Face Database – An Annotated Dataset of 240 Face Images//www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3160 (дата обращения: 03.02.2013).