

А.А. Лукьяница

ЭФФЕКТИВНЫЕ ДЕСКРИПТОРЫ ДЛЯ ЛОКАЛЬНОГО ОПИСАНИЯ ИЗОБРАЖЕНИЙ

Введение

Существует большой класс задач, в которых нужно установить взаимно-однозначное соответствие между точками на двух изображениях, полученных путем регистрации одной и той же сцены либо различными камерами, либо одной камерой при изменении ракурса. Такие задачи возникают, например, при обработке результатов аэрофотосъемки, в фотограмметрии, при сшивке панорамных изображений, при построении трехмерной модели сцены по стереопаре и во многих других областях обработки изображений. Точки, между которыми установлено соответствие, обычно называют сопряженными. Как правило, для решения этой задачи на изображении выделяют так называемые особые (или ключевые) точки (в англоязычной литературе *interest points* или *keypoints*), и находят соответствие между ними, а для остальных точек соответствие устанавливается путем интерполяции. Это, с одной стороны, позволяет существенно сократить объем вычислений, поскольку число ключевых точек существенно меньше, чем всех точек изображения, а с другой стороны – снизить число неправильно найденных соответствий. Последнее обстоятельство связано с тем, что в качестве особых используются те точки изображения, окрестность которых хорошо различима на обоих изображениях и слабо изменяется при изменении освещения, ракурса и масштаба, например, точки, соответствующие вершинам углов, Т-образным соединениям и т.п.

Для последующего сопоставления точек их окрестность описывается вектором характерных признаков, который называется дескриптором (от англ. *descriptor* – описатель, признак). К настоящему моменту предложено большое число алгоритмов для вычисления дескрипторов; среди них наиболее эффективными считаются методы SIFT (*Scale-Invariant Feature Transform*) [1] и SURF (*Speed-Up Robust Features*) [2]. В обоих методах для достижения инвариантности к изменению масштаба строится пирамида изображений, получаемых из исходного путем последовательного уменьшения разрешения, и анализируются изображения всех уровней пирамиды. В методе SIFT к изображению каждого уровня применяется набор гауссовых фильтров

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

с увеличивающимися в $\sqrt{2}$ раз значениями σ , начиная с $\sigma = \sqrt{2}$, а затем вычисляются разности смежных полученных изображений. В качестве особых берутся точки, в которых достигается локальный экстремум этих разностей. В методе SURF для каждой точки изображения строится матрица Гессе, элементами которой являются свертка изображения со вторыми производными гауссиана (1) по каждому направлению:

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{yx}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix}, \quad (2)$$

В качестве особых выбираются точки, в которых достигается локальный максимум определителя этой матрицы.

В методе SIFT дескрипторы вычисляются следующим образом. Выбирается окно – окрестность особой точки размером 8×8 пикселей, и в каждой точке этой окрестности вычисляется градиент, что позволяет избежать от небольших изменений освещенности. Далее градиенты умножаются на весовую гауссову функцию с центром в особой точке и значением σ , равным половине размера окна, что позволяет компенсировать небольшие ошибки в определении положения ключевой точки. Далее строится более грубая сетка размером 4×4 , и значения градиентов для всех пикселей, попавших в укрупненные ячейки, осредняются. После этого для градиентов строятся гистограммы, соответствующие восьми основным направлениям, полученным поворотом горизонтальной оси на угол $\pi/4$. В результате получается вектор характерных признаков, состоящий из $4 \times 4 \times 8 = 128$ элементов. Для снижения влияния возможного изменения освещенности этот вектор нормируется таким образом, чтобы он стал единичной длины. После этого локальное изображение в окрестности особой точки поворачивается в направлении, соответствующем максимальному элементу гистограммы – за счет этого достигается инвариантность к повороту, и дескриптор преобразуется путем выполнения соответствующего поворота [1].

В методе SURF для получения дескрипторов применяется аналогичная техника, только вместо вычисления градиентов используется разложение по вейвлетам Хаара, за счет чего достигается ускорение вычислений. На первом шаге алгоритма находится главное направление путем нахождения максимума суммы коэффициентов разложения в пределах круговых секторов размером $\pi/3$ с центром в особой точке. Далее изображение поворачивается в этом направлении, разлагается по вейвлетам и умножается на гауссову весовую функцию. В качестве дескриптора используется вектор, состоящий из суммы коэффициентов разложения в горизонтальном и вертикальном направлениях, а также из суммы модулей этих коэффициентов [2].

В настоящей работе рассматриваются два типа дескрипторов – в полярных и в полярно-логарифмических координатах. Очевидными достоинствами полярных дескрипторов является легкость их вычисления, а также инвариантность относительно поворота. Второй тип дескрипторов основан на применении полярно-логарифмического преобразования, которое поворот изображения и изменение его масштаба в исходных декартовых координатах переводит в параллельный перенос вдоль главных осей в полярно-логарифмических координатах. Если для сопоставления этих дескрипторов находить максимум Фурье-корреляции, то можно добиться инвариантности относительно поворота и изменения масштаба изображений.

Полярные дескрипторы

Пусть особая точка имеет координаты (x_0, y_0) . Введем локальные полярные координаты (r, φ) с центром в этой точке, как показано на Рис. 1А, и построим в полярных координатах дискретную сетку:

$$r_k = -\frac{\Delta p}{2} + k \cdot \Delta p, \quad k = 1, \dots, N_r,$$

$$\varphi_n = -\frac{\Delta \varphi}{2} + n \cdot \Delta \varphi, \quad n = 1, \dots, N_\varphi, \quad \Delta \varphi = \frac{2\pi}{N_\varphi}.$$

Здесь N_r и N_φ соответственно число точек сетки по радиусу и углу, а Δp – размер пикселя.

Интерполируем значения интенсивностей изображения в узлы этой сетки. Рассмотрим произвольный узел с координатами в полярной системе (r_k, φ_n) . В декартовых координатах он будет иметь координаты

$$x_r = x_0 + r_k \cos \varphi_n, \quad y_r = y_0 + r_k \sin \varphi_n.$$

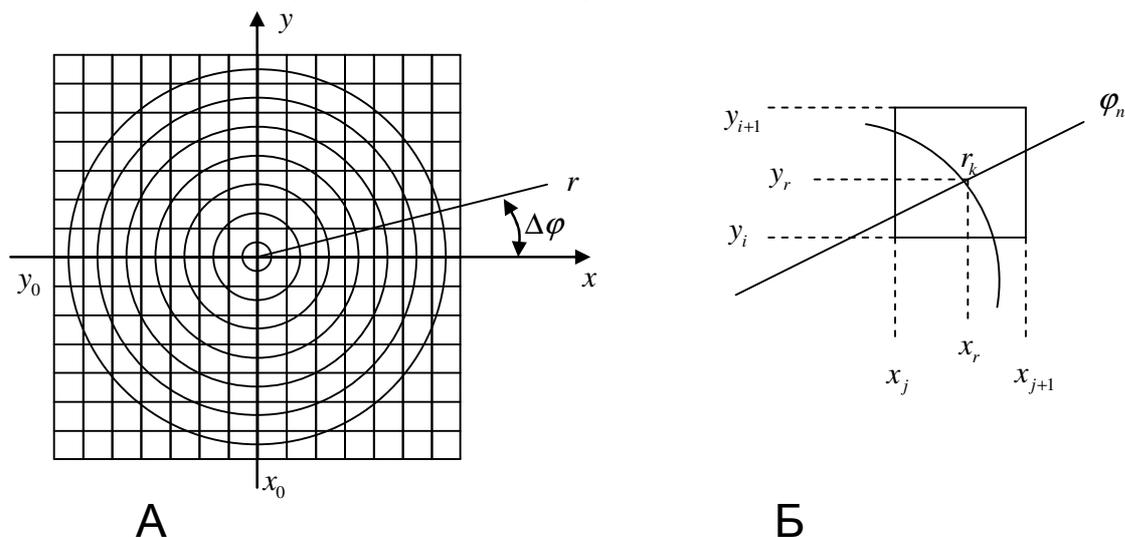


Рис. 1. А – полярные координаты, Б – аппроксимация изображения на полярную сетку.

Найдем такие индексы i и j , при которых одновременно выполняются следующие неравенства (см. Рис. 1Б):

$$x_j \leq x_r \leq x_{j+1}, \quad y_i \leq y_r \leq y_{i+1},$$

и вычислим интерполяционные коэффициенты α и β :

$$\alpha = \frac{x_r - x_j}{x_{j+1} - x_j}, \quad \beta = \frac{y_r - y_i}{y_{i+1} - y_i}.$$

С помощью этих коэффициентов можно провести интерполяцию изображения в узлы полярной сетки:

$$J_{k,n} = (1 - \alpha) \left[(1 - \beta) I_{i,j} + \beta I_{i+1,j} \right] + \alpha \left[(1 - \beta) I_{i,j+1} + \beta I_{i+1,j+1} \right]. \quad (3)$$

Вычислим центральные разностные производные по углу φ :

$$\frac{\partial J(r_k, \varphi_n)}{\partial \varphi} = \begin{cases} J_{k,2} - J_{k,N_\varphi}, & n = 1 \\ J_{k,n+1} - J_{k,n-1}, & 1 < n < N_\varphi \\ J_{k,1} - J_{k,N_\varphi-1}, & n = N_\varphi \end{cases} \quad (4)$$

Вычислим суммарную интенсивность абсолютных величин производных вдоль каждого радиуса:

$$S_n = \sum_{k=1}^{N_r} \left| \frac{\partial J(r_k, \varphi_n)}{\partial \varphi} \right|, \quad n = 1, \dots, N_\varphi,$$

и найдем максимальное значение величины S_n

$$S_{\max} = \max S_n, \quad n = 1, \dots, N_\varphi.$$

И наконец, повернем полярную систему координат таким образом, чтобы нулевое значение угла соответствовало направлению S_{\max} . Это обеспечит инвариантность относительно поворота изображения в окрестности особой точки. В качестве дескриптора будем использовать значения разностных производных (4) в повернутой системе координат.

Полярно-логарифмические дескрипторы

Полярно-логарифмическое (в англоязычной литературе Log-Polar) [3,4] преобразование формально описывается следующими соотношениями:

$$I(x, y) \rightarrow I(\rho, \varphi),$$

$$\rho = \log_b r, \quad \varphi = \arctg \left(\frac{y - y_0}{x - x_0} \right), \quad r = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (5)$$

Здесь (x_0, y_0) – положение особой точки (см. Рис. 2А), а основание логарифма выбирается как $b = \exp \left(\frac{1}{r_{\max}} \ln r_{\max} \right)$, где r_{\max} – максимальное

значение радиуса окружности r , в границах которой рассматривается дескриптор.

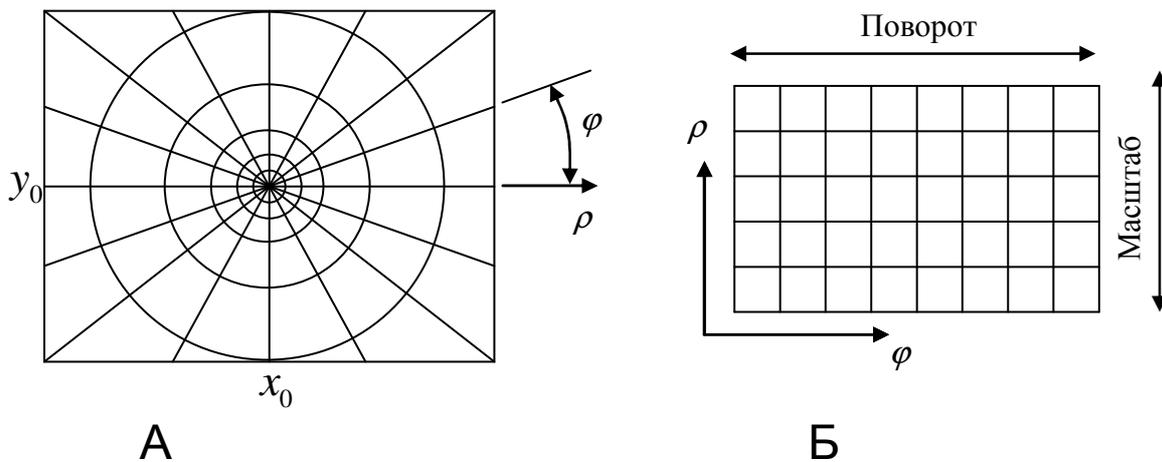


Рис. 2. А – Полярно-логарифмические координаты, Б – поворот и изменение масштаба изображения приводят к параллельному переносу соответственно по осям φ и ρ .

Как видно из формул (5), (r, φ) являются полярными координатами, так что

$$x = r \cos \varphi + x_0, \quad y = r \sin \varphi + y_0, \quad r = b^\rho.$$

Инвариантность полярно-логарифмических преобразований состоит в том, что при повороте изображения (в декартовых координатах) происходит параллельный перенос изображения вдоль оси $O\varphi$ (в полярно-логарифмических координатах), а изменение масштаба приводит к параллельному переносу изображения вдоль оси $O\rho$. Последнее непосредственно следует из свойств логарифма:

$$\log \lambda x = \log \lambda + \log x.$$

На Рис. 2 Б показано, что поворот приводит к сдвигу по оси абсцисс, а изменение масштаба – к сдвигу вдоль оси ординат.

Для построения дескриптора введем равномерные сетки по ρ и φ в полярно-логарифмических координатах и, как и в описанном выше случае с полярной системой координат, интерполируем изображение $I_{i,j}$ в узлы этой сетки, используя формулу (3). К полученному изображению, которое обозначим как $J_{k,n}$, $k = 1, \dots, N_\rho$, $n = 1, \dots, N_\varphi$, применим дискретное преобразование Фурье, которое обозначим как F , и в результате получим дескриптор:

$$D(k, n) = F(J_{k,n}).$$

Этот дескриптор обладает свойствами инвариантности к сдвигу. Действительно, рассмотрим два изображения, одно из которых получено из другого в результате сдвига:

$$J_1(k, n) = J_2(k - k_0, n - n_0). \quad (6)$$

После применения к этому выражению преобразования Фурье получим [5]:

$$F_1(\xi, \eta) = e^{-2\pi i(\xi k_0 + \eta n_0)} F_2(\xi, \eta).$$

Таким образом, при смещении изображения меняется только его фазовый спектр. Это позволяет вычислять близость дескрипторов следующим образом. Сначала нужно вычислить нормализованную кросс-спектральную функцию

$$C = \frac{D_1 D_2^*}{|D_1 D_2|}, \quad (7)$$

где символ * обозначает комплексное сопряжение, а потом применить к этой функции обратное преобразование Фурье:

$$c_{k,n} = F^{-1}(C). \quad (8)$$

Схожесть дескрипторов D_1 и D_2 определяется как величина максимального значения этой функции:

$$d(D_1, D_2) = \max_{k,n} c_{k,n}. \quad (9)$$

Если дескрипторы D_1 и D_2 соответствуют изображениям (6), полученным в результате сдвига, то

$$C = e^{-2\pi i(\xi k_0 + \eta n_0)},$$

поэтому

$$c_{k,n} = \delta(k - k_0, n - n_0),$$

где δ – дельта-функция Дирака. В этом случае максимальное значение величины (8) соответствует положению сдвига (k_0, n_0) .

Таким образом, поскольку полярно–логарифмическое преобразование переводит изменение масштаба и поворота изображения в сдвиги относительно полярно-логарифмических координат, а преобразование Фурье инвариантно относительно сдвигов, то описанные дескрипторы обладают инвариантностью к изменению масштаба и повороту изображения, если для их сравнения использовать формулы (7) – (9).

Численный эксперимент

Для проведения численного эксперимента использовались три пары изображений: стереопара с очень большим параллаксом (Рис. 3 А, Б); стереопара, у которой одно изображение повернуто на 10^0 и уменьшено на 12% (Рис. 3 В, Г) и два последовательных кадра из фильма (Рис. 3 Д, Е).



А



Б



В



Г



Д



Е

Рис. 3. Изображения, на которых проводилось тестирование.

Сравнение проводилось путем использования дескрипторов четырех видов: SIFT, SURF, полярных и полярно логарифмических. Первые два типа дескрипторов были реализованы путем использования графической библиотеки с открытыми кодами OpenCV версии 2.4, для двух других был разработан численный код на алгоритмическом языке С++. Поскольку дескрипторы в методах SIFT и SURF имели 128 элементов, при

использовании полярных и полярно-логарифмических дескрипторов сетка по углу содержала 16 точек, а сетка по радиусу – 8 точек, так что дескрипторы также имели по 128 элементов. Для каждого i -го дескриптора D_i левого изображения находился наименее отличный от него дескриптор правого изображения D_j ; схожесть дескрипторов d_{ij} вычислялась по следующим формулам:

$$d_{ij} = \min_j d(D_i, D_j)$$

– для дескрипторов, вычисленных по методам SIFT, SURF и в полярных координатах, и по формуле

$$d_{ij} = \max_j d(D_i, D_j)$$

– в полярно-логарифмических координатах. Далее находились минимальное d_{\min} и максимальное d_{\max} значения величины d_{ij} , после чего считалось, что дескрипторы соответствуют одной и той же точке на изображениях, если $d_{ij} \leq d_{\min} + \theta(d_{\max} - d_{\min})$ в полярных координатах и для методов SIFT и SURF, а в полярно-логарифмических координатах – если $d_{ij} \geq d_{\max} - \theta(d_{\max} - d_{\min})$. Здесь $0 \leq \theta \leq 1$ - величина порога; чем больше его значение, тем большее число точек попадают в разряд соответствующих друг другу, но при этом также возрастает количество ошибочно отождествленных. Величина порога подбиралась следующим образом: от минимального значения 0.1 он увеличивался до тех пор, пока не появлялась хотя бы одна ошибка в определении сопряженных точек. При обработке указанных изображений значение порога находилось в диапазоне от 0.16 до 0.41. Результаты численного эксперимента приведены в Табл. 2 и Табл. 1.

Табл. 1. Относительное время вычисления дескрипторов

SIFT	SURF	Polar	Log-Polar
1.000	1.042	0.021	0.126

В Табл. 1 находятся относительные затраты процессорного времени для вычисления одного дескриптора по сравнению с методом SIFT, причем в последнем столбце учитывалось также время сравнения двух дескрипторов между собой (поскольку для этого использовалась Фурье-корреляция). Видно, что метод SURF имеет почти такую же вычислительную скорость, как и SIFT; вычисление полярных

дескрипторов происходит на два порядка быстрее, а полярно-логарифмических – на порядок быстрее.

Табл. 2. Количество правильно найденных сопряженных точек относительно метода SIFT

№	SIFT	SURF	Polar	Log-Polar
1	1	0.69	1.15	3.88
2	1	0.94	0.87	3.00
3	1	1.06	1.12	2.57

В Табл. 2 приведены результаты сравнения рассмотренных методов по количеству найденных сопряженных точек. В первом столбце указан номер пары изображений, в остальных столбцах – относительное количество правильно найденных сопряженных точек по сравнению с числом точек, верно отождествленных методом SIFT. Анализ таблицы позволяет сделать следующие выводы. Метод SIFT несколько превосходит метод SURF. Полярные дескрипторы позволяют получить практически такие же результаты, как и в методе SIFT, и уступают только при обработке повернутого и уменьшенного изображения. Поэтому их использование будет оправдано при решении задач, в которых критической является скорость расчетов, поскольку они существенно проще остальных в вычислительном плане. Полярно-логарифмические дескрипторы существенно превосходят все остальные, они позволяют получать почти в три раза больше сопряженных точек за существенно меньшее, почти в десять раз, время.

Заключение

В настоящей работе рассмотрены дескрипторы в полярных и полярно-логарифмических координатах, и проведено их сравнение с наиболее эффективными к настоящему времени дескрипторами, используемыми в методах SIFT и SURF. Проведенный численный эксперимент показал, что использование полярных дескрипторов позволяет получить практически такие же результаты по количеству верно найденных сопряженных точек, как в методах SIFT и SURF, однако за существенно меньшее время. Дескрипторы в полярно-логарифмических координатах позволяют получить почти в три раза больше сопряженных точек, чем при использовании любого из рассмотренных методов, и в восемь раз быстрее, чем по методам SIFT и SURF, что позволяет сделать вывод о существенно превосходстве этого типа дескрипторов над всеми рассмотренными.

Литература

1. Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints // *International Journal of Computer Vision*. 2004. 60. № 2, P. 91–110.
2. Bay H., Ess A., Tuytelaars T., Gool L. V. SURF: Speeded Up Robust Features // *Computer Vision and Image Understanding (CVIU)*. 2008. 110. №3. P. 346–359.
3. Torres-Méndez L.A., Ruiz-Suárez J. C., Sucar L. E., Gómez G. Translation, Rotation, and Scale-Invariant Object Recognition // *IEEE Transactions on Systems, Man and Cybernetics*. 2000. 30, № 1. P. 125–130.
4. Zokai S., Wolberg G. Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations // *IEEE Transactions on Image Processing*. 2005. 14. № 10. P. 1422–1434.
5. Castro E., Morandi C. Registration of Translated and Rotated Images Using Finite Fourier Transforms // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1987. 9. № 5. P. 700–703.