

УДК 004.932

Сравнение методов сопоставления признаков объектов на изображении

Власов Николай Вячеславович

Гудым Антон Васильевич (научный руководитель)

МГТУ им. Н.Э. Баумана, Москва, Россия

Работа содержит в себе содержательное и математическое описание алгоритмов распознавания объектов на изображении. В качестве рассматриваемых алгоритмов представлены такие методы выделения признаков, как SIFT, ROOTSIFT и LIGHTGLUE при различных вариациях, а также рассмотрен процесс создания тестовых данных для их работы. Сравнение методов было осуществлено при помощи метрики Precision и Recall.

Ключевые слова: компьютерное зрение, сравнение алгоритмов, метрика Precision и Recall, распознавание объектов на изображении, матрица гомографии, выделение признаков на изображении

Введение. Распознавание объектов на изображении — это ключевая задача в области компьютерного зрения, которая включает в себя обнаружение и классификацию объектов на изображении. Существует множество методов для решения этой задачи включая как традиционные методы обработки изображений, так и методы на основе машинного обучения и нейронных сетях.

Актуальность обоснована множеством сфер применения, от медицины до аграрной отрасли.

Рассматриваются алгоритмы, которые для своей работы используют ключевые точки изображения (keypoints). Для нахождения ключевых точек на изображениях и последующего их сравнения используются три составляющие:

- детектор или алгоритм поиска точек;
- дескриптор или алгоритм описания точек (descriptor);
- матчер или алгоритм сопоставления и фильтрации (matcher).

Исследуем следующие алгоритмы поиска признаков: SIFT, ROOTSIFT, LIGHTGLUE (ROOTSIFT), LIGHTGLUE (SUPERPOINT), LIGHTGLUE (DISK),

Масштабно-инвариантная трансформация признаков (SIFT) является алгоритмом выявления признаков в компьютерном зрении для нахождения и описания локальных признаков в изображениях. Впервые он был представлен Д. Лоу в 1999 году [1]. Алгоритм состоит из четырех основных шагов:

- определение экстремумов в масштабном пространстве;
- локализации ключевых точек;
- назначение ориентации;
- построение дескриптора ключевой точки.

ROOTSIFT это метод сопоставления, который за свою основу берет ранее описанный метод SIFT. Метрика расстояния, используемая в SIFT, — это евклидово расстояние, которое представляет собой расстояние между двумя пикселями по прямой линии. Существуют и другие метрики расстояний, такие как Манхэттенское расстояние, шахматная доска, квазиевклидовы расстояния, расстояние Чебышева, расстояние Хеллингера и т. д. Вместо евклидова расстояния для SIFT, RootSIFT использует расстояние Хеллингера [2]

$$H(x, y) = \sum_{i=1}^n \sqrt{x_i y_i},$$

где x, y — нормализованные гистограммы; n — их количество.

Метод LightGlue представляет собой усовершенствованный подход к сопоставлению ключевых точек в изображениях, который стремится улучшить точность и эффективность по сравнению с традиционными методами, такими как SIFT (ScaleInvariant Feature Transform) и SuperGlue [3]. LightGlue сочетает в себе преимущества глубокого обучения и классических методов компьютерного зрения для достижения высокой производительности при сопоставлении ключевых точек.

Метрики сравнения алгоритмов Precision и Recall. Для сравнения алгоритмов использовалась метрика Precision и Recall. Precision (точность) и Recall (полнота) — это метрики, используемые для оценки производительности классификационных моделей. Они помогают понять, насколько хорошо модель определяет положительные классы [4].

Precision — определяет параметр точности модели, т. е. соотношение правильно найденных пар соответствий, к общему количеству найденных соответствий, включая неправильные. Precision определяется как

$$\text{Precision} = \frac{TP}{TP+FP}.$$

Recall — определяет параметр полноты модели, т. е. соотношение правильно найденных пар соответствий, ко всем существующим правильных соответствий. Recall определяется формулой

$$\text{Recall} = \frac{TP}{TP+FN},$$

где TP — истинно положительный результат; FP — ошибочно положительный результат; TN — истинно негативный результат; FN — ошибочно негативный результат.

Исходные данные. В качестве тестовых данных для сопоставления было выбрано изображение, представленное на рис. 1 слева, которое затем было подвергнуто преобразованию при помощи матрицы гомографии, а именно повороту относительно осей x, y и z на произвольные углы. Это сделано для того, чтобы при сравнении алгоритмов учитывалось влияние вращения и перспективы. В общей сложности было создано 125 изображений. Эти изображения сопоставляются с другим набором данных, содержащим 9 изображений с различным содержанием и положением исходного изображения.

В результате получается 1125 сравнений. Пример сравнения представлен на рис. 1, где слева представлено исходное изображение, а справа рассматриваемое, зеленым цветом обозначены сопоставленные ключевые точки.



Рис. 1. Пример сопоставления изображений

Матрица гомографии. Матрица гомографии, также известная как матрица проективного преобразования, используется для описания проективного преобразования между двумя плоскостями [5]. Она широко применяется в компьютерном зрении и фотограмметрии для сопоставления изображений, коррекции перспективы и других задачах, поскольку хорошо описывает геометрические искажения информации при ее проекции на экран (в небольшой окрестности ключевой точки). Матрица гомографии является 3×3 матрицей, которая включает в себя следующие элементы:

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}. \quad (1)$$

Здесь каждый элемент h_{ij} определяет параметры преобразования, такие как вращение, масштабирование, сдвиг и перспективные искажения.

Сама матрица определяется как $H = H_S H_A H_P$, где H_S отвечает за преобразование масштаба и сдвига, т. е. сохраняет углы и пропорции; H_A отвечает за аффинные преобразования, т. е. искажает углы, но сохраняет параллельность; H_P отвечает за перспективное искажение, т. е. параллельные линии сходятся в точку.

Алгоритм сравнения. Для сравнения методов распознавания объектов на изображении был реализован следующий алгоритм:

1) нахождение ключевых точек (keypoints) и текстурных признаков (descriptors) соответствующим методом;

2) определение матрицы гомографии (1) для исходного изображения при помощи функции `findHomography` библиотеки OpenCV;

3) создание эталонного множества `groundtruth` путем применения матрицы гомографии к исходному изображению;

4) сравнение множества `groundtruth` с множеством соответствий `matches` при разных фильтрующих коэффициентах, используя метрику Precision и Recall.

Результаты сравнения алгоритмов. В результате сравнения алгоритмов метрикой Precision и Recall был получен следующий график, представленный на рис. 2, где численные значения — это количество найденных пар соответствий ключевых точек.

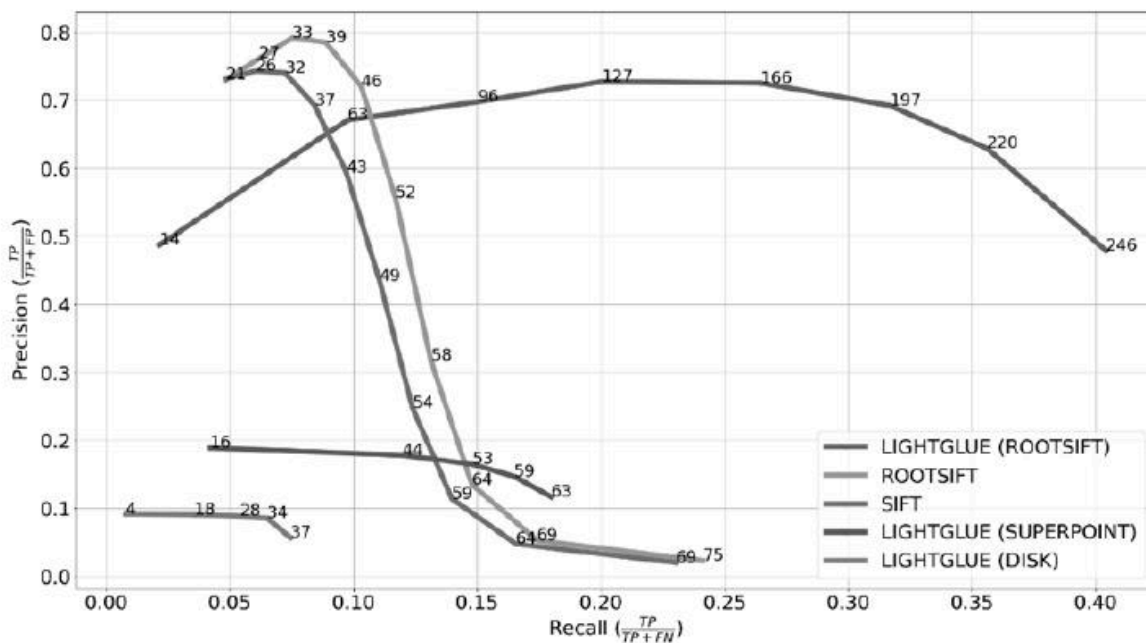


Рис. 2. Результат сравнения алгоритмов метрикой Precision и Recall

По представленному графику можно сделать вывод о том, что наибольшей точностью обладает метод ROOTSIFT, хотя и имеет при этом низкий показатель Recall, т. е. полноты. Метод, продемонстрировавший наибольшее значение Recall, а также максимальное число найденных соответствий оказался LIGHTGLUE (ROOTSIFT), что говорит о лучшем алгоритме фильтрации и поиска.

В зависимости от задачи, могут быть важны разные метрики: точность (precision), полнота (recall). Для биометрической аутентификации или диагностика заболеваний по медицинским изображениям важна высокая точность (precision), для сегментации объектов или формирования карты окружения необходима полнота и объем данных (recall).

Для анализа результатов также рассматривались площади кривых (рис. 2), чем большую площадь покрывают кривые, тем лучшим считается их результат. Сравнение происходило относительно кривой SIFT в областях их перекрытия с той или иной кривой. Результат анализа приведен в таблице.

Результат сравнение площадей под кривыми графика Precision и Recall

№	Метод	Площадь под исследуемой кривой (AUC)	Площадь под кривой SIFT (AUC)	Отношение площадей (относительно SIFT)
1	ROOTSIFT	0,065	0,053	1,22
2	LIGHTGLUE (ROOTSIFT)	0,147	0,053	2,77
3	LIGHTGLUE (SUPERPOINT)	0,031	0,092	0,33
4	LIGHTGLUE (DISK)	0,004	0,043	0,09

По представленной таблице видно, что наилучшие результаты дает метод LIGHTGLUE (ROOTSIFT), после которого идет ROOTSIFT. Все остальные методы показали результат хуже, относительно SIFT.

Заключение. В результате работы были реализованы следующие алгоритмы поиска признаков: SIFT, ROOTSIFT, LIGHTGLUE (ROOTSIFT), LIGHTGLUE (SUPERPOINT), LIGHTGLUE (DISK), а также проведено их сравнение при помощи метрики Precision и Recall. Полученные результаты показали, что ROOTSIFT имеет превосходство над стандартным SIFT и его использование позволяет получать лучший результат. LIGHTGLUE, основанным на ROOTSIFT, имеет несомненное преимущество по количеству правильно найденных соответствий, а соотношение precision к recall говорит о более точном алгоритме фильтрации.

Проведенные вычислительные эксперименты показали, что скорость выполнения сопоставлений у LIGHTGLUE различается примерно в два раза, при 1 часе у SIFT и ROOTSIFT, и 2 часа у LIGHTGLUE. При исходных данных в виде 10000 ключевых точек и 1125 итераций сопоставлений (GPU: NVIDIA GeForce RTX 3070; CPU: Intel Core i5-11400F; RAM: 32gb).

Литература

- [1] Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints, 2004, pp 5–25. URL: <https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf> (accessed 21.08.2024).
- [2] Zisserman A., Arandjelovi R. Three things everyone should know to improve object retrieval. Department of Engineering Science, 2012, pp. 2–7. URL: <https://www.robots.ox.ac.uk/~vgg/publications/2012/Arandjelovic12/arandjelovic12.pdf> (accessed 21.08.2024).
- [3] Lindenberger P., Sarlin P., Pollefeys M. ETH Zurich, Microsoft Mixed Reality and AI Lab. LightGlue: Local Feature Matching at Light Speed, 2023, pp. 3–8. URL:

https://openaccess.thecvf.com/content/ICCV2023/papers/Lindenberger_LightGlue_Local_Feature_Matching_at_Light_Speed_ICCV_2023_paper.pdf (accessed 21.08.2024).

- [4] Tharwat A. Classification assessment methods. Faculty of Computer Science and Engineering. Frankfurt, Frankfurt University of Applied Sciences, 2021, pp. 169–176.
<https://doi.org/10.1016/j.aci.2018.08.003>
- [5] Hartley R. Multiple View Geometry in Computer Vision Second Edition. Cambridge, Cambridge University Press, 2004, pp. 34–45.

Comparison of methods for comparing features of objects in an image

Vlasov Nikolay Vyacheslavovich

Gudym Anton Vasilievich (*scientific advisor*)

BMSTU, Moscow, Russia

The paper contains a substantial and mathematical description of algo-rhythms for recognizing objects in an image. As the considered algorithms, feature extraction methods such as SIFT, ROOTSIFT and LIGHTGLUE under different variations are presented and the process of creating test data for their operation is discussed. The comparison of the methods in this paper was done using the Precision and Recall metrics.

Keywords: computer vision, algorithm comparison, Precision and Recall metrics, recognizing objects in an image, homography matrix, feature extraction in an image