

SIFT на практике.

В погоне за понимаем мощных современных нейронок давайте сделаем неспешную паузу и обратим внимание на базовый метод из классического CV.

SIFT – Scale Invariant Feature Transform – хороший инструмент для извлечения локальных признаков из изображения (почти как свертка!), который, как следует из названия, еще и инвариантен к некоторым аффинным преобразованиям, изменению яркости, контрастности и зашумленности [2]. Почему и как – попробуем разобраться на примерах.

В основе SIFT лежит идея о нахождении и вычислении характеристик особых точек на изображении, лежащих в местах наибольшего изменения контрастности. Для их нахождения к исходному изображению пошагово применяют гауссовский фильтр (Gaussian Blur) и вычисляют разницу между получившимися сглаживаниями (см. Рис. 1, DoG: Difference of Gaussians). Данная процедура повторяется для изображения в разных масштабах. После этого для каждого пикселя проверяется, является ли он экстремумом:

- 1) Среди своих 8 соседей на текущем изображении.
- 2) На изображениях того же масштаба, но с другой DoG.
- 3) Если проверки 1 и 2 пройдены, то координаты пикселя пересчитываются для определения его положения на отмасштабированных изображениях.

Проверки 1-3 повторяются для всех масштабов изображения, называемых октавами.

Обычно большое количество точек находятся на границе объектов, однако в данном подходе для большей стабильности метода они удаляются (Рис. 2). Также удаляются точки, чья интенсивность не превышает определенного порога.

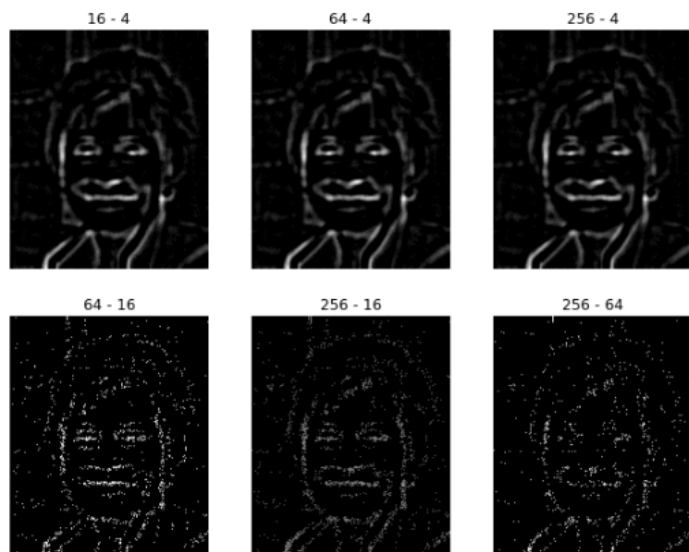


Рис. 1. Разность изображений с разным параметром сглаживания sigma (разности гауссианов).

После того, как ключевые точки и их координаты и масштаб определены, для них вычисляется ориентация градиента интенсивности на основе интенсивности близлежащих пикселей. По набору этих характеристик создается дескриптор:

- 1) Берутся соседние пиксели окном 16x16.
- 2) Это пространство делится на квадраты 4x4. Всего 16 квадратов.
- 3) Для каждого квадрата строится гистограмма ориентаций пикселей по 8 направлениям (С, Ю, З, В, СЗ, СВ, ЮЗ, ЮВ).
- 4) Вся полученная информация собирается в вектор длиной $16 \times 8 = 128$ признаков. Это и есть дескриптор, при помощи которого можно сравнивать ключевые точки на разных изображениях!

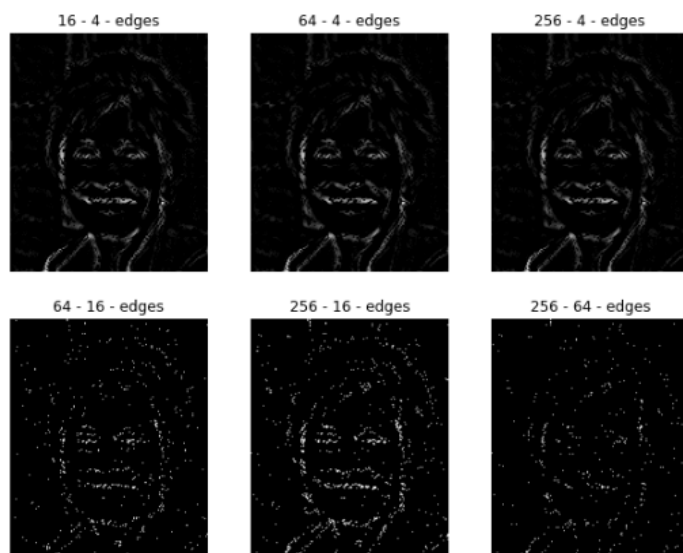


Рис. 2. Удаление границ изображения из разностей гауссианов.

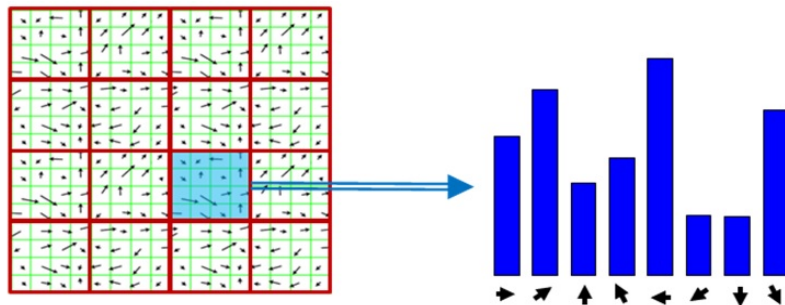


Рис. 3. Схема вычисления дескриптора. [3].

Попробуем подход на практике. Возьмем несколько аугментаций (повороты, масштабирование, несколько видов Blur, Sharpen, зеркальные отражения, шум, изменение контраста и яркости) изображения Judge Judy и посмотрим, какие из них будут иметь больше совпадений локальных признаков. Для сравнения ключевых точек используем евклидову метрику и две реализации SIFT: из OpenCV и из scikit-image.

На рисунке 4 представлены некоторые результаты эксперимента для OpenCV. Для scikit-image результаты сопоставимы.

Сравнивая количество найденных совпадений и дистанции между дескрипторами ключевых точек на разных изображениях, можно сделать вывод, что SIFT действительно хорошо справляется с поворотами, сопоставлением по части изображения (когда после ShiftScaleRotate доступно не все изображение полностью), масштабированием, добавлением небольшого шума, небольшим изменением яркости и контрастности. Однако, этот метод менее полезен в случаях зеркального отображения, транспонирования, сильного шума, сильного изменения яркости, контраста или сглаживания.

Также бросается в глаза, что обычные опорные точки (landmarks) лица выступают ключевыми точками и для SIFT детектора.



Рис. 4. Результаты работы OpenCV SIFT детектора на аугментированных изображениях в сравнении с исходным. А) Поворот на 180 градусов. В) Изменение яркости (предел -0.1) и контраста (предел 0.4). С) Изменение яркости (предел -0.4) и контраста (предел -0.4). D) Sharpen (коэффициент 0.8). Е) Гауссовский шум (дисперсия 4000). F) Горизонтальное зеркально отображение. G) Поворот на -28 градусов со сдвигом 0.25 и масштабированием 0.3. H) Изменение яркости (предел 0.4) и контраста (предел 0.1). I) Размытие (коэффициент 25).

Что посмотреть по теме:

- 1) Ноутбук с экспериментами из этого поста (https://nbviewer.org/github/ul611/SIFT_over_augmentations/blob/main/SIFT.ipynb).
- 2) Статьи изобретателя метода D. Lowe (<https://www.cs.ubc.ca/~lowe/papers/iccv99.pdf>, <https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf>).
- 3) Статья про вычисление дескрипторов (<https://gilscvblog.com/2013/08/18/a-short-introduction-to-descriptors/>)

Что попробовать дальше:

- 1) Попробовать взять другие изображения Джуди и протестировать на них сопоставление изображений с помощью SIFT. Если совпадений будет мало, либо они будут неверными, не отчаивайтесь, это нормально.