

## **Применение метода гауссовых пирамид и алгоритма SIFT-дескрипторов в задаче распознавания языка жестов**

А.П. Матросов

Московский физико-технический институт (государственный университет)

Алгоритм SIFT-дескрипторов (Scale-Invariant Feature Transform) [1] был создан Дэвидом Дж. Лоуи (David G. Lowe) в 1999 году как действенный метод поиска инвариантных ключевых точек для сопоставления изображений. В данной работе он находит применение в решении задачи распознавания предопределенных классов изображений на примере чтения языка жестов.

В терминах SIFT ключевой точкой объекта является такая точка, которая с большой долей вероятности будет найдена на любом изображении этого объекта. Детектор, извлекающий эти точки, должен обеспечивать инвариантность их нахождения относительно преобразований смещения, поворота, изменения масштаба, яркости, световых условий и положения наблюдателя. SIFT-дескриптор - это идентификатор ключевой точки, полученный с помощью алгоритма SIFT.

В основе алгоритма лежит построение пирамиды гауссианов и разностей гауссианов. Под гауссианом понимается изображение, преобразованное гауссовским ядром, т.е. размытое по гауссу, а разностью гауссианов - изображение, полученное путем попиксельного вычитания двух гауссианов с различными радиусами размытия. На первом шаге алгоритма строится пирамида гауссианов исследуемого изображения, с разделением по октавам. Затем в полученном масштабируемом пространстве ищутся особые точки, за которые принимается любой локальный экстремум разности гауссианов - т.е. любой пиксель, значение разницы гауссианов в котором больше или меньше 26 значений в окружающих его пикселях. На следующих этапах происходит выделение ключевых точек из особых. Для начала координаты особой точки определяются с субпиксельной точностью с помощью аппроксимации функции разностей гауссианов многочленом

Тейлора. Затем точка проходит проверку на контраст, близость к границе объекта и плохую освещенность, в конечном счете получая ранг ключевой. Далее для ключевой точки вычисляется ее ориентация, как функция направлений градиентов точек, соседствующих с ней. На финальном этапе алгоритма строится дескриптор точки путем светки данных об ориентации гауссовым ядром и упорядочивания их в векторы гистограмм. Данный алгоритм был детально изучен Брайаном Расселом в задаче распознавании образов в больших массивах данных [2].

Для распознавания языка жестов используется взвешенная вероятностная модель. Для начала алгоритм обучается на изолированных примерах азбуки Дактиля, тем самым формируя набор особых точек, которые им соответствуют, и вероятность их проявления. Затем его работа тестируется в реальных условиях, где стоит задача обнаружить и распознать символы в произвольно данном изображении. Задача считается успешно выполненной, если алгоритм способен обнаружить и распознать 80% символов в повседневных изображениях среднего качества.

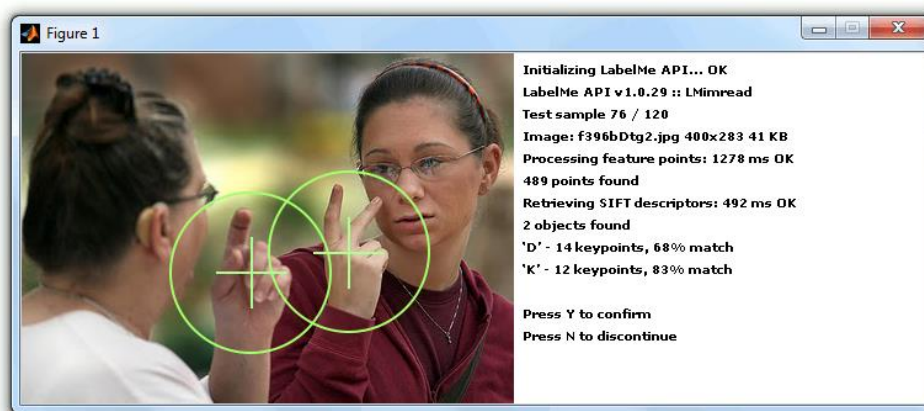


Рис. 1. Обученный алгоритм в действии - обнаружение символов D и K

## Литература

1. *David G. Lowe*. Distinctive Image Features from Scale-Invariant Keypoints // CSD, University of British Columbia. 2004.
2. *Br. Russell*. Labeling, Discovering, and Detecting Objects in Images. // MIT. 2008.