# Распознавание объектов по локальным масштабно-инвариантным признакам

## Аннотация

Разработана система распознавания объектов, использующая новый класс локальных признаков изображения. Признаки инвариантны к масштабированию, перемещению и вращению изображения и частично инвариантны к изменениям освещения и аффинной или трехмерной проекции. Эти признаки имеют сходные свойства с нейронами в нижней височной коре, которые используются для распознавания объектов в зрении приматов. (Эти признаки имеют сходные свойства с нейронами в нижней височной коре приматов, которые используются для распознавания объектов) Признаки эффективно обнаруживаются с помощью поэтапной фильтрации, которая идентифицирует стабильные точки в пространстве масштаба. Создаются ключи изображения, которые допускают локальные геометрические деформации, представляя размытые градиенты изображения в нескольких плоскостях ориентации и в нескольких масштабах. Ключи используются в качестве входных данных для метода индексации ближайшего соседа, который идентифицирует совпадения объектов-кандидатов. Финальная проверка каждого совпадения достигается путём нахождения решения методом наименьших квадратов для неизвестных параметров модели. Результаты экспериментов показывают, что надежное распознавание объектов может быть достигнуто на беспорядочных частично закрытых изображениях со временем вычисления менее 2 секунд.

## Введение

Распознавание объектов в загроможденных сценах реального мира требует наличия локальных признаков изображения, на которые не влияют близлежащие помехи или частичная окклюзия. Признаки должны быть хотя бы частично инвариантны к освещению, трехмерным преобразованиям и обычным вариациям объектов. С другой стороны, признаки также должны быть достаточно отличительными, чтобы идентифицировать конкретные объекты среди множества альтернатив. Сложность задачи распознавания объектов во многом связана с отсутствием успеха в нахождении таких признаков изображения. Однако недавнее исследование использования плотных локальных признаков (например, Schmid & Mohr [19]) показало, что эффективное распознавание часто может быть достигнуто за счет использования локальных дескрипторов изображений, отобранных в большом количестве повторяющихся мест.

В этой статье представлен новый метод генерации признаков изображения, который называется масштабно-инвариантной трансформацией признаков (SIFT *scale-invariant feature transform*). Этот подход преобразует изображение в большой набор векторов локальных признаков, каждый из которых инвариантен к перемещению, масштабированию и вращению изображения и частично инвариантен к изменениям освещения и аффинной или трехмерной проекции. Предыдущие подходы к генерации локальных признаков не обладали инвариантностью к масштабу и были более чувствительны к проективным искажениям и изменениям освещения. Признаки SIFT имеют ряд общих свойств с реакциями нейронов в нижней височной (IT) коре головного мозга приматов. В этом документе также описываются улучшенные подходы к индексации и верификации моделей.

Масштабно-инвариантные признаки эффективно идентифицируются с помощью поэтапной фильтрации. На первом этапе идентифицируются ключевые местоположения в масштабном пространстве путем поиска местоположений, которые являются максимумами или минимумами функции разности Гаусса. Каждая точка используется для генерации вектора признаков, который описывает локальную область изображения, выбранную относительно ее системы координат масштабного пространства. Признаки достигают частичной инвариантности к локальным вариациям, таким как аффинные или трехмерные проекции, за счет размытия мест градиента изображения. Этот подход основан на модели поведения сложных клеток коры головного мозга зрения млекопитающих. Полученные векторы признаков называются ключами SIFT. В текущей реализации каждое изображение генерирует порядка 1000 ключей SIFT, что требует менее 1 секунды вычислений.

Ключи SIFT, полученные из изображения, используются в методе индексации ближайшего соседа для идентификации моделей объектов-кандидатов. Наборы ключей, которые согласуются с потенциальной позицией модели, сначала идентифицируются с помощью хеш-таблицы преобразования Хафа, а затем с помощью метода наименьших квадратов для окончательной оценки параметров модели. Когда по крайней мере 3 ключа согласуются с параметрами модели с минимизированным остатком, есть убедительные доказательства присутствия объекта. Поскольку на изображении типичного объекта могут быть десятки ключей SIFT, можно иметь значительные уровни окклюзии на изображении и при этом сохранять высокий уровень надежности.

Текущие объектные модели представлены в виде двумерных местоположений ключей SIFT, которые могут подвергаться аффинной проекции. Достаточные вариации в расположении признаков позволяют распознавать перспективную проекцию плоских форм при повороте до 60 градусов от камеры или допускать поворот 3D-объекта до 20 градусов.

## Связанные исследования

Распознавание объектов широко используется в индустрии машинного зрения для проверки, регистрации и манипулирования. Однако современные коммерческие системы распознавания объектов зависят почти исключительно от сопоставления шаблонов на основе корреляции. Хотя это очень эффективно для определенных инженерных сред, где положение объекта и освещение жестко контролируются, сопоставление шаблонов становится невыполнимым с вычислительной точки зрения, когда вращение объекта, масштаб, освещение и положение в пространстве могут варьироваться, и тем более при работе с частичной видимостью и большой моделью. базы данных.

Альтернативой поиску совпадений во всех местах изображения является извлечение признаков из изображения, которые хотя бы частично инвариантны для процесса формирования изображения и соответствуют только этим признакам. Было предложено и исследовано множество типов признаков, включая линейные сегменты [6], группы ребер [11, 14] и регионы [2], среди многих других предложений. Хотя эти функции хорошо работают для определенных классов объектов, они часто не обнаруживаются достаточно часто или с достаточной стабильностью, чтобы сформировать основу для надежного распознавания.

Недавно была проведена работа по разработке гораздо более плотных коллекций признаков изображения. Один из подходов заключался в использовании углового детектора (точнее, детектора пиков в локальном варианте изображения) для определения повторяющихся областей изображения, вокруг которых можно измерить локальные свойства изображения. Чжан и др. [23] использовали угловой детектор Харриса для определения местоположения признаков для эпиполярного выравнивания изображений, снятых с разных точек зрения. Вместо того, чтобы пытаться сопоставить области одного изображения со всеми возможными областями второго изображения, большая экономия времени вычислений была достигнута за счет сопоставления только областей, центрированных в угловых точках каждого изображения.