

Анализ эффективности метода Bag-of-words в задаче визуальной геолокализации

Кишкун А.А., магистр кафедры безопасности информационных систем
ФГАОУ ВО ГУАП, kichkun@vu.spb.ru

Веселов А.И., доцент, кандидат технических наук кафедры
инфокоммуникационных систем ФГАОУ ВО ГУАП, felix@vu.spb.ru

Аннотация.

В данной статье рассматривается способ решения задачи визуальной геолокализации по фотографии методом Bag-of-words, оценивается возможность применения данного метода для решения такого рода задач.

Введение

Навигаторы и путеводители в мобильных телефонах стали неотъемлемой частью жизни. Одной из актуальных задач в данной области является задача визуальной геолокализации – определения местоположения по фотографии. Одним из актуальных вопросов в данной задаче является вопрос классификации объектов городской инфраструктуры - зданий. Задача классификации зданий в условиях городской среды является важным шагом в различных приложениях, таких как автоматические системы наведения, обнаружение слежки, движение роботов в режиме реального времени, его локализация и визуальная навигация, проектирование, дизайн, 3D-реконструкции города, а также гиды по городу. В данной статье будет рассмотрен способ решения данной задачи методом Bag-of-words, будет исследована возможность применения метода [1] для задачи классификации зданий.

Постановка задачи

Для постановки задачи введем несколько определений: будем называть *тренировочной выборкой* базу изображений, на которой производится обучение алгоритма; *запросным изображением* – изображение, поступающее на вход обученного алгоритма для классификации, *эталонным изображением* – изображение из

тренировочной выборки, относящиеся к тому же классу, что и верно классифицированное запросное. В данной задаче *классом* является адрес объекта, а *объектами* являются здания. *Особой, или ключевой, точкой (feature point)* m называют точку изображения, окрестность которой $o(m)$ можно отличить от окрестности любой другой точки изображения $o(n)$ в некоторой другой окрестности особой точки $o(m)$ [2]. В качестве окрестности точки изображения для большинства алгоритмов берётся прямоугольное окно размером 5x5 пикселей. Процесс определения особых точек достигается путём использования детектора и дескриптора. *Детектор* – метод извлечения особых точек из изображения. Детектор обеспечивает инвариантность нахождения одних и тех же особых точек относительно преобразований изображений. *Дескрипторы* – идентификаторы особой точки, выделяющие её из остального множества особых точек. Существует ряд алгоритмов для построения дескрипторов, таких как SIFT – Scale Invariant Features Transform [3], SURF – Speed Up Robust Features [4] и др. В основе построения дескрипторов зачастую лежит гистограмма ориентированных градиентов (HOG) [5], которая используется для определения ориентации особой точки. Дескриптор строится на основе направлений всех градиентов окрестности.

Таким образом, задачу визуальной геолокализации можно свести к задаче классификации, в которой конечным множеством объектов является множество адресов зданий, представленных на запросных изображениях, а множеством входных объектов являются здания, представленные изображениями в тренировочной выборке. В данной статье будет рассмотрен способ решения данной задачи методом Bag-of-words.

Этапы алгоритма

На рисунке 1 представлена общая схема предлагаемого решения.

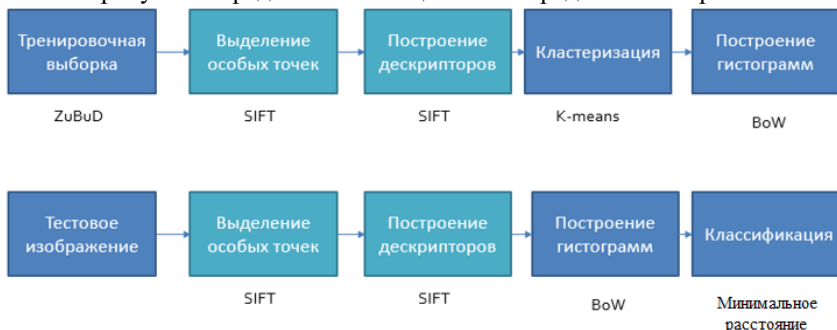


Рисунок 1: Общая схема предлагаемого решения.

В качестве тренировочной выборки использовалась база данных изображений Zurich Buildings Database for Image Based Recognition (ZuBuD) [6], в которой каждый класс представлен пятью изображениями, отличающимися углом съемки. Пример изображений выборки ZuBuD представлен на рисунке 2.



Рисунок 2: Пример изображений тренировочной выборки ZuBuD

Алгоритм состоит в следующем: каждый из n классов тренировочной выборки представляется k изображениями (для тренировочной выборки ZuBuD $k=5$). Далее выделяются особые точки и строятся их дескрипторы. В качестве детекторов и дескрипторов особых точек использовался метод SIFT с количеством точек m . Таким образом, каждый класс представлен $m*k$ дескрипторами, где каждый дескриптор x представлен 128 числами. Общее количество извлеченной на данном этапе (см. Рис. 3) информации равно $n*k*m*128=n*312500$ (расчет для 500 особых точек на изображение, количество особых точек ограничено не было и составляло от 500 до 2500).



Рисунок 3: Извлечение особых точек, построение дескрипторов.

Извлеченная на предыдущем этапе информация, является входной для алгоритма Линде-Бузо-Грея (k-means) [7]. Суть алгоритма заключается в следующем: на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения центра масс кластеров. Это происходит за конечное число итераций, так как количество возможных разбиений конечного множества конечно, а на каждом шаге суммарное квадратичное отклонение V не увеличивается, поэтому заикливание невозможно. Таким образом, все извлеченные на предыдущем этапе дескрипторы разбиваются на p наборов центроидов, по которым строятся гистограммы (см. Рис. 4). Было сделано предположение, что дескрипторы, извлеченные с k изображений одного класса имеют более близкие значения, чем дескрипторы изображений другого класса. Таким образом алгоритм Линде-Бузо-Грея позволит получить описатель класса, однозначно характеризующий данный класс. Результаты эксперимента представлены на рисунках 5 и 6. Представленные гистограммы визуальны различимы, более формально представленное различие будет рассмотрено ниже. Необходимо отметить, что алгоритм Линде-Бузо-Грея является локально-оптимальным: в зависимости от случайно-выбранных начальных условий алгоритм сходится к разным решениям. В рамках данного исследования планируется изучение влияния данной особенности алгоритма на результат.

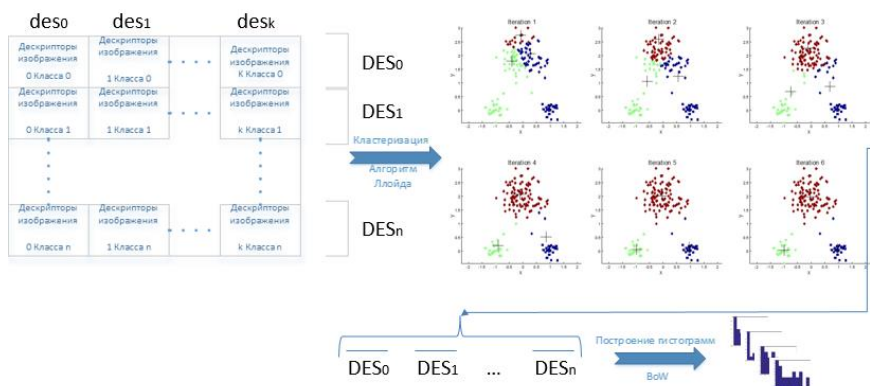


Рисунок 4: Применение алгоритма Линде-Бузо-Грея, построение гистограмм.

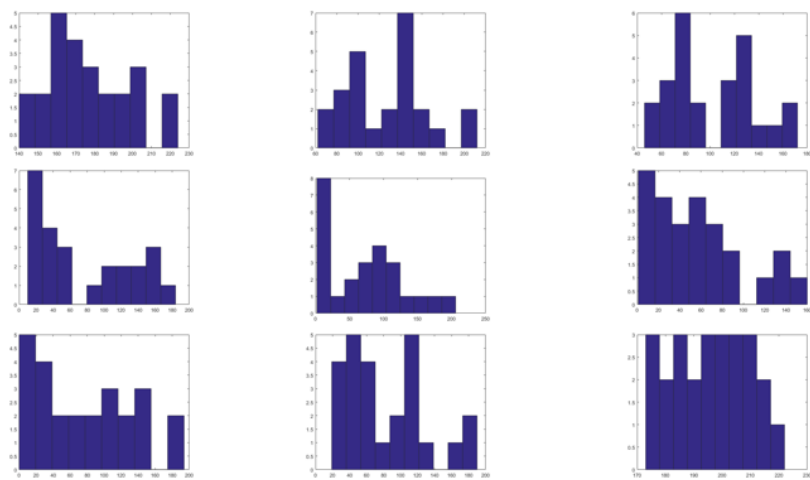


Рисунок 5: Результат построения гистограмм для 9 классов по 5 изображениям на класс.

Результаты

Для оценки эффективности алгоритма была проведена классификация квантованных дескрипторов, извлеченных с запросного изображения. Тренировочная выборка была разбита на две части: обучающую, содержащую 4 изображения для каждого класса и тестовую,

содержащую 1 новое изображение для каждого класса. На стадии обучения обучающая часть использовалась для построения гистограмм: методом SIFT были построены дескрипторы, построена кодовая книга, выполнено квантование, построены эмпирические плотности вероятностей квантованных дескрипторов для каждого класса. Далее было проведено тестирование запросными изображениями, для каждого из которых были выделены дескрипторы, проквантованы по кодовой книге, построенной на обучении, найдена функция плотности вероятностей и выбрана наиболее близкая к ней из ранее обученных. Исследование проводилось для выборки из 10, 20, 30, 40 и 50 классов. Для 10 классов такой классификатор на каждом отложенном изображении выдает правильную метку, для 20 классов ошибка классификации составляет 20%, для 30 – 13,3%, для 40 - 25%, для 50 – 16% (см. Рис. 6а). Было проведено исследование зависимости результата классификации от параметров алгоритма Линде-Бузо-Грея: зависимость результата от максимального количества итераций не была выявлена (см. Рис. 6б). Моделирование проводилось при помощи пакета VLFeat [8] в среде MatLab.

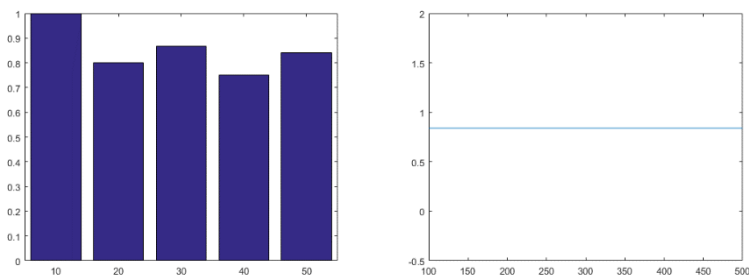


Рисунок 6а: Зависимость вероятности верной классификации от размера выборки

Рисунок 6б: Зависимость ошибки классификации от количества итерация

Заключение

Проведя исследование эффективности применения методом Bag-of-words для решения задач визуальной геолокализации зданий, можно сделать следующие выводы: для выбранного сценария метод показывает хорошие результаты. На выборке из 10 классов (5 изображений на класс) результат классификации составляет 100%.

При увеличении выборки вероятность верной классификации снижается, что обусловлено недостаточным количеством экземпляров класса при обучении ($k=4$), а также сходством изображений различных классов. На результат классификации также влияют выбранные параметры

алгоритмов, такие как максимальное количество итераций алгоритма k-means, количество извлекаемых дескрипторов, алгоритм выбора метки класса и другие. В рамках данного исследования была выдвинута гипотеза о зависимости точности классификации от параметров алгоритма Линде-Бузо-Грея. Данная гипотеза не подтвердилась.

Проведенный анализ позволяет сделать вывод об эффективности данного метода для решения поставленной задачи. В дальнейших исследованиях планируется рассмотреть эффективность данного метода на выборках большего размера, а также улучшить результат классификации, исследовав параметры алгоритмов.

Список литературы

1. Two bag-of-words classifiers // MIT computer science and artificial intelligence laboratory URL: <http://people.csail.mit.edu/fergus/icc2005/bagwords.html> (дата обращения: 20.05.2017).
2. Конушин А. С. Слежение за точечными особенностями сцены (Point feature tracking) //Компьютерная графика и мультимедиа. Электрон. журн. – 2003. – №. 1. – С. 5.
3. Ng P. C., Henikoff S. SIFT: Predicting amino acid changes that affect protein function //Nucleic acids research. – 2003. – Т. 31. – №. 13. – С. 3812-3814.
4. Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features //Computer vision–ECCV 2006. – 2006. – С. 404-417.
5. Tsai G. Histogram of oriented gradients //University of Michigan. – 2010.
6. Shao H., Svoboda T., Van Gool L. Zubud-zurich buildings database for image based recognition //Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland, Tech. Rep. – 2003. – Т. 260. – С. 20.
7. Ostrovsky R. et al. The effectiveness of Lloyd-type methods for the k-means problem //Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on. – IEEE, 2006. – С. 165-176.
8. VLFeat URL: <http://www.vlfeat.org/> (дата обращения: 20.05.2017).