

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Математико-механический факультет

Кафедра информационно-аналитических систем

Зареченский Михаил Алексеевич

Обнаружение текста в коллекциях
изображений, содержащих текст на разных
языках

Бакалаврская работа

Допущена к защите.
Зав. кафедрой:
д. ф.-м. н., профессор Новиков Б. А.

Научный руководитель:
к. ф.-м. н. Васильева Н. С.

Рецензент:
ст. преп. Рецензент А. И.

Санкт-Петербург
2015

SAINT-PETERSBURG STATE UNIVERSITY
Mathematics & Mechanics Faculty

Chair of The Analytical Information Systems

Mikhail Zarechenskiy

Text detection in image collections with multilingual text

Graduation Thesis

Admitted for defence.
Head of the chair:
professor Boris Novikov

Scientific supervisor:
Natalia Vassilieva

Reviewer:
assistant assistant

Saint-Petersburg
2015

Оглавление

Введение	4
Постановка задачи	5
1. Обзор существующих подходов	6
1.1. Методы, использующие априорное знание о структуре текста	6
1.2. Методы, основанные на глубоком обучении	7
Заключение	8
Список литературы	9

Введение

Последние достижения в области цифровых технологий позволяют делать снимки практически с любого мобильного устройства. Как результат, количество фотографий, сделанных пользователями, растёт каждый день. В тоже время, коллекции с большим количеством изображений очень долго аннотировать вручную и поэтому они часто остаются без каких-либо сведений, кроме аннотаций, сделанных самим устройством.

Наличие текста на изображении часто даёт важную информацию о семантике изображения, как, например, название магазина, улицы, название предмета на изображении и так далее. Проаннотированные изображения можно использовать для различных задач анализа изображений, таких как поиск изображений, автоматическая навигация. Другой популярной идеей является моментальный перевод текста на изображении с мобильного устройства. Снятая фотография, содержащая иностранный язык, обрабатывается на предмет обнаружения и распознавания текста для дальнейшего перевода на требуемый язык.

Такие сценарии использования говорят о том, что изображения часто могут быть получены с мобильного устройства и иметь низкое разрешение, размытие, шум. Более того, часто заранее не может быть известен язык, на котором написан текст на изображении.

Актуальной исследовательской задачей является эффективное обнаружение и распознавание текста на изображениях. Обнаружение текста является одним из самых важных шагов при дальнейшем распознавании текста. В данной работе будет исследоваться именно задача обнаружения текста.

Важной особенностью данной работы является рассмотрение задачи обнаружения текста по отношению к изображениям, содержащим текст на разных языках. Будет показано, что эффективность многих существующих подходов к обнаружению текста может сильно зависеть от языка, представленного на изображении. Для того, чтобы оценить инвариантность тех или иных методов к языку, будет рассмотрена теоретическая основа методов, а также будут поставлены эксперименты с различными данными, как содержащими только текст на английском языке, так и с данными, содержащими текст на разных языках.

Постановка задачи

В рамках дипломной работы были поставлены следующие задачи:

- Проанализировать наиболее эффективные методы обнаружения текста
- Описать и выделить основные этапы работы алгоритмов предметной области
- Аналитически определить зависимость каждого из этапов алгоритмов обнаружения текста к языку
- Поставить ряд экспериментов на данных, как содержащих только английский текст, так и на данных, содержащих текст на разных языках
- Провести эмпирический анализ и сделать выводы о зависимости алгоритмов предметной области к языку

1. Обзор существующих подходов

Для наиболее эффективного распознавания текста, первым шагом текст на изображении должен быть точно обнаружен, однако это является довольно сложной задачей в связи с большой вариацией представления текста на изображении. Текст может иметь различные вариации шрифта, стиля, размера, искажения, иметь различную контрастность из-за разных условий освещения. Всё изображение так же может сильно варьироваться, следует учитывать низкое разрешение, низкую контрастность, неоднородный фон. Такое разнообразие порождает различные подходы к обнаружению текста.

С одной стороны, априорное знание о структуре текста, привело к развитию алгоритмов, которые можно поделить на три категории: алгоритмы, основанные на анализе текстуры, алгоритмы на основе компонент связности [9], [2], [8], [6] и гибридные методы [5].

С другой стороны, методы, основанные на глубоком обучении (deep learning), использующие небольшое количество априорной информации о данных, получили широкое распространение в таких областях задач анализа изображений, как классификация и распознавание объектов. В том числе, существуют и алгоритмы обнаружения текста, работающие на основе свёрточных нейронных сетей [1], [7].

Далее рассмотрим подробнее алгоритмы обнаружения текста из каждого класса.

1.1. Методы, использующие априорное знание о структуре текста

К алгоритмам, использующим априорное знание о структуре текста относят алгоритмы, основанные на анализе текстуры, алгоритмы на основе компонент связности и гибридные методы.

Алгоритмы, основанные на анализе текстуры, выделяют текстурные особенности из изображения, после чего классификатор идентифицирует наличие текста. Как правило, текстурные особенности выделяются с использованием техники скользящего окна. Данная процедура довольно трудоёмка, и, более того, необходимо провести данную операцию для разных масштабов. Алгоритмы, основанные на данном подходе имеют тенденцию к низкой производительности и значительно теряют точность при падении качества изображений.

В тоже время алгоритмы, основанные на связных компонентах, извлекают из изображений отдельные регионы - кандидаты в символы. Далее применяются различные эвристики для фильтрации выделенных компонент. Оставшиеся компоненты(регионы) группируются в текст. Компоненты группируются в текст как правило двумя способами: на основе геометрических особенностей и при помощи методов кластеризации.

зации. На последнем шаге также возможна дополнительная фильтрация для устранения ошибок второго рода.

Гибридные методы на первом шаге производят сегментацию изображения таким образом, чтобы символы, расположенные на изображении, имели отчётливую границу с фоном. Далее выделяются компоненты связности и применяются различные эвристики для исключения компонент, не представляющих символы. На последнем шаге компоненты группируются в текст.

Стоит отметить, что в соответствии с результатами соревнования "Multi-script Robust Reading Competition" [4] на конференции ICDAR 2013, подход, основанный на связных компонентах, показал себя, как наиболее эффективный и производительный по сравнению с остальными подходами. Поэтому, далее данный подход будет рассматриваться более подробно.

1.2. Методы, основанные на глубоком обучении

Методы, основанные на глубоком обучении, позволяют отойти от выбора вручную признаков для обнаружения текста. Признаки извлекаются из коллекции изображений, предназначенной для обучения. При этом вид признаков зависит от метода обучения. Например, при использовании свёрточных нейронных сетей [3], признаки ограничиваются свёрточными масками и их комбинациями.

Заключение

Список литературы

- [1] End-to-End Text Recognition with Convolutional Neural Networks / T. Wang, D. J. Wu, A. Coates, A.Y. Ng // IEEE International Conference on Pattern Recognition (ICPR). — 2012.
- [2] Gomez L., Karatzas D. Multi-script Text Extraction from Natural Scenes // ICDAR. — 2013.
- [3] Gradient-Based Learning Applied to Document Recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // IEEE. — 1998.
- [4] Kumar Deepak, Prasad M. N. Anil, Ramakrishnan A. G. Multi-script robust reading competition in ICDAR 2013 // ACM - Proc. International Workshop on Multilingual OCR. — 2013.
- [5] Milyaev S., Barinova O., Novikova T. Image binarization for end-to-end text understanding in natural images // ICDAR. — 2013.
- [6] Robust text detection in natural images with edge-enhanced maximally stable extremal regions / H. Chen, S.S. Tsai, G.Schroth и др. // IEEE International Conference on Image Processing. — Sep 2011.
- [7] Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning / A. Coates, B. Carpenter, C. Case et al. // ICDAR. — 2011.
- [8] Yalniz I. Zeki, Gray Douglas, Manhmatha R. Adaptive exploration of text regions in natural scene images // ICDAR. — 2013.
- [9] Yin X.-C., Yin X., Huang K. Robust text detection in natural scene images // CoRR. — 2013. — T. abs/1301.2628.