УДК 004.932

ЛОКАЛИЗАЦИЯ ТЕКСТА НА ИЗОБРАЖЕНИЯХ СЛОЖНЫХ ГРАФИЧЕСКИХ СЦЕН

Андрианов А. И.

ООО «Аби ИнфоПоиск», Москва, Россия (127273, Москва, улица Отрадная, 2Б, стр.6, офис 7-02), e-mail: Andrew_An@abbyy.com

Проведен анализ существующих методов поиска и локализации текста на изображениях. Определены требования к системе распознавания текста на изображении сложных графических сцен, характеризующиеся неоднородностью фона, отсутствием чётких критериев отличия фона от текста, большой вероятностью разнообразных искажений. Определена возможность использования для решения задачи выделения связных областей изображения метода MSER. Для классификации областей изображения на «текстовые» или «не-текстовые» предлагается использовать классификаторы AdaBoost, SVM. В качестве самостоятельного признака возможно использование алгоритма SWT. Он может использоваться как при обучении классификаторов, так и при объединении букв-кандидатов в слова или строки текста.

Ключевые слова: методы распознавания, методы локализации текста, сложные графические сцены.

TEXT AREA DETECTION ON COMPLICATED IMAGES

Andrianov A. I.

OOO «ABBYY InfoPoisk», Moscow, Russia (127273, Moscow,Otradnaya street, 2B, building 6,office 7-02), e-mail: Andrew_An@abbyy.com

The methods of detection and locating the text areas on complicated images were researched. The requirements for the text recognition system for images contained complicated graphic scenes were determined. These images can have inhomogeneous background with no clear criteria's to distinguish text and background. In addition these images can have various distortions. We have researched the applicability of MSER's method for the solving the problem of selection the connected image areas. We propose to use AdaBoost's and SVM's classifiers for separation of the regions of the image as "textual" and "non-textual" segments. We use SWT algorithm as an additional attribute. It can be used for classifiers training, as well as for merging of probable-letters to words or to the lines of a text.

Key words: recognition methods, text detection methods, images contained complicated graphic scenes.

Введение

Задача распознавания текста на изображениях сложных графических сцен подразумевает локализацию и распознавание текста в ситуации, когда изображения кроме текста могут содержать большое количество другой графической информации, которая практически не поддается фильтрации. Сложные графические сцены – это изображения, характеризующиеся неоднородностью фона, отсутствием чётких критериев отличия фона от текста, большой вероятностью разнообразных искажений. Примером сложных графических сцен могут являться кадры из кинофильмов с субтитрами, фотографии уличных вывесок, сделанные цифровым фотоаппаратом и т.д.

Существующие на данный момент системы распознавания текста ориентированы на бумажные документы, то есть на типовые условия, в которых однотонный текст расположен на однотонном контрастном фоне. Такие изображения могут содержать нетекстовую

информацию, например, рисунки, графики. Однако эта информация легко локализуется, отделяется от текстовых областей.

В отличие от распознавания изображений бумажных документов, распознавание текста на сложных графических сценах осложняется тем, что текст на таких изображениях часто не отделен от прочей информации явно, а является частью этой информации. Невозможно заранее предугадать, в какой области изображения расположен текст, какое он имеет искажение. Соответственно, усложняется задача поиска, локализации текста.

Целью исследования является разработка и исследование методов и алгоритмов поиска и локализации текста на фотографиях и видеокадрах, позволяющих улучшить показатели точности выделения текста в ситуации сложных графических сцен.

Частичные решения задачи поиска текста на изображениях сложных графических сцен уже описаны в научно-технической литературе. Например, широко известны системы для распознавания государственных номерных знаков автомобилей на изображениях, полученных с видеокадров камер слежения [1]. Решение этой задачи облегчается тем, что количество форматов номерных знаков ограничено, размер и положение номерного знака относительно автомобиля фиксированы, а текст знака всегда находится на контрастном фоне.

Другой пример частного решения поставленной задачи – распознавание метки даты и времени на сканированном изображении фотографии [2]. Эта задача подразумевает лишь поиск цифр в определенной области изображения.

Решение задачи локализации текста в сложных графических сценах в общем виде позволит расширить область практического применения систем распознавания текста. Одним из наиболее интересных вариантов применения являются так называемые системы «дополненной реальности». Возможность распознавать текст с видоискателя фотоаппарата мобильного устройства позволит реализовать перевод по наведению. При наведении видоискателя на сцену, содержащую текст на иностранном языке, пользователь получает на видоискателе изображение, в котором этот текст подменен текстом на языке пользователя.

Другой пример использования полного решения обсуждаемой задачи – оцифровка массива фотографий для дальнейшего текстового поиска.

На основании предполагаемых областей применения можно сформулировать требования к системе распознавания текста на изображении сложных графических сцен.

1. Скорость. Необходимо иметь возможность использовать систему на современных мобильных устройствах (смартфонах);
2. Помехоустойчивость. Искажение, поворот, изменение освещенности не должны влиять на процесс локализации и распознавания текста;

3. Поддержка различных языков распознавания.

Классически задача распознавания текста делится на три фазы:

1. Анализ изображения и выделение текстовых областей;
2. Распознавание символов;
3. Сборка текста и экспорт данных во внешний формат.

Как упоминалось выше, основное отличие обсуждаемой задачи от существующих решений заключается в первой фазе – нужно найти текст в сложной графической сцене. Проведем обзор методов, предлагаемых в научно-технической литературе для решения этой задачи.

Работа [3] описывает идею автоматической генерации признаков, используемых для распознавания. Основная идея здесь заключается в том, что признаки для классификатора можно не составлять вручную, а создавать их с помощью машинного обучения. В качестве базы для обучения предлагается использовать так называемые патчи, имеющие размер 8x8 пикселов. Дальнейшая обработка изображения, по мнению авторов, может быть сведена к вычислению обученных признаков на интересующих областях изображения.

Основная идея ряда работ (например, [4]) заключается в том, что буквы и слова на изображении, как правило, имеют постоянную толщину штриха. Поэтому для выявления таких объектов, по мнению авторов, перспективно использовать алгоритм SWT. Действительно, разумно ожидать, что ширина штриха в букве должна быть приблизительно постоянна, равно как и буквы в одном слове должны иметь похожую ширину. Признак тем более полезен, так как использовать его возможно как в качестве одного из признаков при классификации регионов, так и в качестве признака при «собирании» регионов в слова.

Границы символов в рамках описанного подхода могут вычисляться, например, с помощью Canny Egde Detector. Затем в направлении градиента находится парная граничная точка. Если градиент в этой точке примерно коллинеарен градиенту в первой точке, то все точки между ними заполняются значением найденной ширины. Затем, при втором проходе по найденным на первом проходе отрезкам вычисленное значение ширины ограничивается медианным значением вдоль отрезка. При этом следует помнить, что предложенный метод потребует определённых дополнительных вычислительных затрат для борьбы с ошибками на углах и некоторыми другими специфическими для него эффектами.

Работа Y. Kunishige и соавторов [5] опирается на идею использования «контекста окружения» (environmental context). Основная мысль заключается в использовании информации о том, что окружает область-«кандидата». Иными словами, предлагается анализировать тот фон, на котором находится регион изображения, возможно, являющийся текстовым. Идея базируется на эмпирическом предположении, что вероятность наличия текста, например, на травяном покрове или на небе – низка.

Второй материал того же автора [6], а также статья авторского коллектива Lao, Du и др. [7] посвящены разработке идеи так называемых feature-точек.

Авторский коллектив [6] предлагает сначала детектировать на изображении так называемые SURF-точки. Предполагается, что если на исследуемом изображении присутствуют буквы, то они будут плотно такими точками покрыты. Дополнительно к этому вычисляется visual saliency (визуальная заметность). Вместе SURF и saliency будут представлять собой (128 + 1) – мерный вектор признаков. На этом векторе предполагается провести обучение классификатора AdaBoost.

В работе [7] к аналогичной задаче предложен несколько иной подход. В общих чертах там тоже решается задача поиска точечного текста. Но для обнаружения точек, составляющих буквы, применяется достаточно хорошо известный алгоритм FAST. Затем производится эвристическая фильтрация ложных кандидатов, объединение точек в буквы, букв – в слова, после чего применяется классификатор SVM (см. [8]) для детектирования текстовых областей.

С одной стороны, все перечисленные работы обещают хорошие результаты при поиске текстов, изначально состоящих из обособленных точек (например, такими текстами набираются цифры, указывающие срок годности на упаковке продуктов). С другой стороны, на сложных графических сценах, где вполне возможно присутствие «не-точечного» текста (рекламные щиты, вывески магазинов, автомобильные номера и т. д.), предложенные авторами алгоритмы дадут не слишком хорошие результаты – именно в силу изначальной своей «ориентированности» на работу с точками.

Оба упомянутых классификатора – AdaBoost и SVM – представляют собой хорошо известные решения, которые будут весьма полезны для решения поставленной задачи.

В работе [9] рассмотрена идея суперпикселов. Сам по себе предлагаемый алгоритм – специфичный для данной задачи, практически не используемый в других решениях – выглядит следующим образом.

Сначала выполняется кластеризация, в процессе которой пикселы изображения собираются в т. н. суперпикселы. На этих своеобразных кластерах производится вычисление признаков, предварительно обученных именно на такую работу, т. е. с суперпикселами. Затем выполняется классификация «текст» или «не текст», для чего используется упомянутый ранее алгоритм AdaBoost. По результатам строится CRF (2-мерный граф суперпикселов), учитывающий как результаты классификатора, так и значения извлечённых признаков. Параметры CRF тоже обучаются на заранее размеченной базе изображений.

Таким образом, предлагаемый в [9] подход отличается высокой сложностью при сравнимом качестве результатов. Можно со всей уверенностью признать эту идею неподходящей для решения обсуждаемой задачи.

Идея MSER-регионов (Maximally stable extremal regions), описанная в [10], [11], напротив, производит весьма благоприятное впечатление. Метод выделения связных компонент, используемый авторами, известен достаточно давно. Суть его в том, что на изображении выполняется бинаризация со всеми порогами. В результате удаётся устойчиво различать регионы, которые при описанном преобразовании мало меняются в некотором заданном интервале порогов. Эти регионы и являются MSER.

Достаточно близка по смыслу к [10] статья [12]. Авторы, также прорабатывая различные варианты применения MSER, нашли специальный способ отсекать неподходящие варианты ещё на стадии их построения. Несколько иная, но близкая по смыслу возможность исследована в [13]. В ней предлагается способ ускоренного (по отношению к ранее упомянутым способам) построения MSER.

В работах [14], [15] предлагается использование габоровских признаков для классификации регионов. С точки зрения качества распознавания предложенные признаки, вероятно, должны продемонстрировать хороший результат. Однако процесс вычисления габоровских признаков будет связан с серьёзными вычислительными затратами. Иными словами, такой алгоритм обещает быть довольно медленным и не удовлетворять требованиям по скорости.

По результатам теоретических исследований для решения задачи локализации текста на изображении сложных графических сцен предлагается использовать метод MSER для выделения связных областей изображения. Для классификации областей изображения, как «текстовых» или «не-текстовых», предлагается использование классификаторов AdaBoost и/или SVM.

Отдельно стоит отметить алгоритм SWT. Он может использоваться в качестве самостоятельного признака как при обучении классификаторов, так и при объединении букв-кандидатов в слова или строки текста.

Работы проводятся при финансовой поддержке Министерства образования и науки Российской Федерации в рамках выполнения ГК 07.514.11.4158 по теме: “Разработка алгоритмов поиска и локализации текста на фотографиях и видеокадрах”.

Список литературы

1. Веснин Е. Н., Царев В. А. Оптимизация процесса обработки данных в системах распознавания буквенно-цифровых меток движущихся объектов // Интеллектуальные

системы и компьютерные науки: Материалы IX международной конференции. – М.: Изд-во механико-математического факультета МГУ, 2006. – Т.2, Ч.1. – С. 73-79.2.

2. A. Shahab, F. Shafait, A. Dengel, Bayesian Approach to Photo Time-Stamp Recognition. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 1039 – 1043, ISBN 978-0-7695-4520-2.

3. A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. Wu, A. Ng, Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 440 - 445, ISBN 978-0-7695-4520-2.

4. B. Epshtein, E. Ofek, Y. Wexler, Detecting Text in Natural Scenes with Stroke Width Transform - 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol.V. San Francisco, 2010.

5. Y. Kunishige, F. Yaokai, S. Uchida, Scenery Character Detection with Environmental Context. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 1049 – 1053, ISBN 978-0-7695-4520-2.

6. S. Uchida, Y. Shigeyoshi, Y. Kunishige, F. Yaokai, A Keypoint-Based Approach Toward Scenery Character Detection. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 819 – 823, ISBN 978-0-7695-4520-2.

7. Y. Du, H. Ai, S. Lao, Dot Text Detection Based on FAST Points. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 435 – 439, ISBN 978-0-7695-4520-2.

8. C. Jung, Q.F. Liu, J. Kim, “Accurate text localization in images based on SVM output scores,” Image and Vision Computing, vol:27, 2009, pp. 1295-1301.

9. M. Cho, J. Seok, S. Lee, J. Kim, Scene Text Extraction by Superpixel CRFs Combining Multiple Character Features - The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 1034 - 1038, ISBN 978-0-7695-4520-2.

10. C. Merino-Gracia, K. Lenc, M. Mirmehdi, A Head-mounted Device for Recognizing Text in Natural Scenes - The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, ISBN 978-0-7695-4520-2.

11. M. Donoser and H. Bischof, “Efficient maximally stable extremal region (MSER) tracking,” in CVPR, 2006, pp. 553–560.

12. L. Neumann, J. Matas, Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 687 - 691, ISBN 978-0-7695-4520-2.

13. J. Zhang and R. Kasturi, “Character energy and link energybased text extraction in scene images,” in ACCV 2010, ser. LNCS 6495, vol. II, November 2010, pp. 832–844.
14. C. Yi, Y. Tian, Text Detection in Natural Scene Images by Stroke Gabor Words. – The 11th International Conference on Document Analysis and Recognition (ICDAR), 2011, Pages: 177–181, ISBN 978-0-7695-4520-2.
15. Q. Liu, C. Jung, and Y. Moon, “Text Segmentation based on Stroke Filter,” Proceedings of International Conference on Multimedia, pp.129-132, 2006.

Рецензенты:

Кузнецов Сергей Олегович, доктор физико-математических наук, профессор, заведующий кафедрой анализа данных и искусственного интеллекта Национального исследовательского университета Высшая школа экономики (НИУ ВШЭ), г. Москва.

Гриненко Михаил Михайлович, доктор физико-математических наук, научный консультант, ООО «Аби ИнфоПоиск», г. Москва.