

Секция «Вычислительная математика и кибернетика»

Распознавание текста на фотографиях с учетом высокоуровневой информации

Новикова Татьяна Владимировна

Студент

Московский государственный университет имени М.В. Ломоносова, Факультет вычислительной математики и кибернетики, Химки, Россия

E-mail: miryable@gmail.com

В последнее время задача распознавания текста на изображениях становится все более актуальной. Примерами приложений являются: системы помощи людям с ограниченными возможностями, аннотация фотографий в интернете, поиск по изображениям, автоматический перевод текста на иностранном языке на фотографиях с мобильных устройств. Существуют алгоритмы, способные достаточно хорошо распознавать текст на отсканированных документах, однако они показывают плохие результаты на более сложных данных, к примеру, на фотографиях.

Задача распознавания текста включает в себя два этапа: поиск текста и распознавание текста. Большинство существующих методов [1, 2, 3] представляют собой системы с последовательной интеграцией стадий обнаружения и распознавания текста. Такой подход, однако, не позволяет использовать одновременно всю имеющуюся информацию о тексте (как низкоуровневую, такую как расположение и распознавание отдельных символов, так и высокоуровневую информацию о внешнем виде текста в целом и лексических особенностях языка).

Предложенный подход к решению задачи рассматривает проблему распознавания текста как проблему совместной оптимизации следующих составляющих: оценка внешних характеристик слова в целом (таких как ориентация, цвет, размер) и соответствия им внешних характеристик отдельных символов, распознавание отдельных символов, соответствие распознавания отдельных слов языковой модели (словарю), моделирование взаимного расположения отдельных символов друг относительно друга. Оптимизация производится с помощью аппарата взвешенных конечных преобразователей (Weighted Finite-State Transducers), который долгое время успешно применяется в задачах обработки текста. Он предоставляет удобное представление контекста распознавания текста и языковых моделей, а также эффективные алгоритмы детерминизации и поиска наилучшего пути, что позволяет эффективно использовать данный аппарат для сформулированной задачи оптимизации.

Было произведено сравнение предложенного метода в контексте задачи распознавания текста на изображениях с одиночными словами, которое показало качественное преимущество предложенного метода в сравнении с доступными результатами других методов [1, 2, 3] распознавания текста.

Литература

1. L. Neumann, J. Matas. "A method for text localization and recognition in real-world images" //In: Proceedings of the 10th Asian conference on Computer vision (ACCV 2010). Queenstown, New Zealand, 2010, Vol. 3, pp. 770–783.

2. K. Wang, B. Babenko, S. Belongie. “End-to-end scene text recognition” //In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV 2011). Barcelona, Spain, 2011, pp. 1457–1464.
3. K. Wang, S. Belongie. “Word spotting in the wild” //In: Proceedings of the 11th European Conference on Computer Vision (ECCV 2010). Crete, Greece, 2010, Vol. 1, pp. 591–694.