УДК 004.93

СПЕЦИАЛИЗИРОВАННЫЙ МЕТОД РАСПОЗНАВАНИЯ ТЕКСТА ДЛЯ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ПАСПОРТНЫХ ДАННЫХ

П. С. Другов, младший научный сотрудник СПИИРАН, магистрант СПбГЭТУ, Санкт-Петербург, Россия;
orcid.org/0000-0002-0319-0554, e-mail: pah53k@gmail.com
Е. Е. Усина, младший научный сотрудник СПИИРАН, магистрант СПбГУАП, Санкт-Петербург, Россия;
orcid.org/0000-0001-9745-0216, e-mail: lizzzi96@mail.ru

При разработке интеллектуального пространства для внедрения на базе имеющейся инфраструктуры предприятия актуальной является задача добавления паспортных данных о новых посетителях в систему предприятия.

***Целью данной статьи** является разработка метода для распознавания текста страниц паспорта гражданина РФ с изображений, полученных с применением камеры высокого разрешения. Разработанный метод ориентирован на работу с паспортом гражданина РФ и включает такие операции, как: увеличение изображения, сегментация объекта (паспорта), обработка изображения фильтрами и распознавание текста. Для тестирования данного метода был разработан вертикальный стенд с областью для размещения документа в верхней части и написан код на языке Python. В результате тестирования программы было выявлено, что самый высокий процент распознавания текста достигается при расстоянии от камеры до документа в 25 см – 88,8 %, тогда, как при увеличении фокусировки камеры и отдалении, были получены результаты 66,7 % и 22,2 % соответственно. Внедрение данного метода на предприятии позволило бы избежать ошибок при переносе информации, обусловленных человеческим фактором (невнимательностью или утомленностью работника), а также сократило бы время занесения данных в клиентскую базу.*

***Ключевые слова:** распознавание текста, OCR, Tesseract, оптическое распознавание символов, OpenCV, обработка изображения, паспорт РФ, интеллектуальное пространство.*

DOI: 10.21667/1995-4565-2019-70-43-51

Введение

Несмотря на достижения вычислительных технологий, при решении задачи распознавания текста остаются актуальными такие проблемы, как: разнообразие форм начертания символов, вариации размеров, масштаба и искажение символов [1]. Данные особенности характерны и для документов государственного образца. Паспорт гражданина РФ содержит разные варианты написания букв в зависимости от года выдачи, что затрудняет программное распознавание текста в таком документе. С целью защиты сведений о личности главная страница паспорта гражданина РФ ламинируется плёнкой с голографическим изображением, что создает дополнительные блики на фотографии и осложняет распознавание текста. Некоторые методы по распознаванию текста на документах являются частью проприетарного программного обеспечения и соответственно не подходят для свободного использования. Этим объясняется актуальность разработки нового метода для распознавания текста на изображении с камеры в потоковом режиме. В ходе представленного исследования разработан метод для перевода данных с паспорта гражданина РФ в текстовый формат. Преимуществами метода являются: обработка изображения документа, защищенного плёнкой с голограммой, обработка бликов на изображении, сегментация страниц и распознавание текста, формат шрифта которого не подлежит свободному распространению.

Обзор существующих методов оптического распознавания текстовых символов

Задача оптического распознавания символов – OCR (Optical Character Recognition) рассматривается множеством научных деятелей и является одной из наиболее распространенных в области компьютерного зрения. В связи с этим существует множество различных подходов для её решения. Условно подходы можно разделить на строчные [2, 3], ориентированные на работу с печатным текстом, и блочные, направленные на распознавание деформированного текста (изогнутые вывески, необычные шрифты и т.д.) [4, 5].

Авторами [2] рассматривается строчный распознаватель текста с открытым исходным кодом, сочетающий в себе глубокие нейронные сети – DNN (Deep Neural Networks) и сети с долгой краткосрочной памятью – LSTM (Long short-term memory), разработанные с использованием библиотеки PyTorch и использующие инструмент CUDA (Compute Unified Device Architecture) для ускорения работы. Суть такого подхода заключается в использовании нейронной сети LSTM и CTC (коннекционная временная классификация [6]) с обучением на последнем слое. Для геометрического выравнивания изображения относительно горизонта сначала проводятся оценка базы (пространственной ориентации строки текста) и оценка высоты символов по оси X. Далее определяется осевая линия, проходящая по центру текстовой строки, и соответственно, осуществляется поворот изображения на угол отклонения осевой линии от горизонтальной оси. Выходные данные сверточного слоя подвергаются 2 видам трансформации: конкатенации и сжатию. Все выходные данные нейронной сети переводятся в вектор класса для каждого состояния и обрабатываются функцией softmax. На стандартном компьютере с графическим процессором метод распознает около 100 строк текста в секунду. По проведенным тестам было доказано, что глубокие гибридные сети с нормализацией геометрических текстовых линий превосходят все другие сети. Еще один пример строчных распознавателей описан в статье [3]. Одним из вариантов применения данного алгоритма является распознавание текста на изображении, полученном с помощью камеры смартфона. На вход подается изображение, которое обрезается до границ слова, подлежащего распознаванию, после чего происходит предварительная обработка изображения, где для фильтрации шума используется медианный фильтр, а для получения черно-белого изображения – алгоритм Otsu [7]. Для распознавания символов используется алгоритм SOM (Self-Organizing Map algorithm) [8], особенностью которого является распознавание одиночных символов, т.е. на этапе предварительной обработки необходима сегментация каждого символа в отдельности. В результате тестирования алгоритм SOM успешно распознал 89,09 % кириллических символов с машинно-генерируемых изображений и 88,89 % кириллических символов на изображениях с камеры телефона. Из 350 фотографий разных слов SOM удалось распознать 292 слова, что составляет 83,42 %. Остальные слова были частично распознаны.

К блочным распознавателям относятся методы, которые способны распознать деформированный текст. В статье [10] авторами предложена новая архитектура нейронной сети, которая совмещает в себе извлечение признаков, моделирование последовательностей и транскрипцию. Ключевыми особенностями данной системы являются: сквозное обучение (end-to-end), отсутствие сегментации символов и привязки к определенному лексикону, генерация небольшой, но эффективной модели. Разработанная нейронная сеть объединяет преимущества как сверточных, так и рекуррентных нейронных сетей. Также авторами была реализована возможность обработки изображений различной длины и обучения сети без подробного описания каждого элемента.

В статье [10] авторы предлагают новую модель распознавания текста с помощью разработанной глубокой нейронной сети – RARE (Robust text recognizer with Automatic REctification), которая приспособлена к работе с неравномерным текстом. RARE состоит из двух сетей: STN (Spatial Transformer Network) и SRN (Sequence Recognition Network). RARE основана на сквозном обучении, поэтому на вход сети требуется подать только изображение и связанные с ним текстовые метки. Сначала текст на изображении геометрически выравнивается относительно горизонта с помощью предсказанных трансформацией сплайнов – это

повышает качество слов. Затем выполняется распознавание текста путем анализа последовательности символов. Эффективность распознавания текстов сквозным обучением все еще сильно уступает по результативности другим методам [1].

В статье [11] авторы описывают процесс предобработки изображения для алгоритма распознавания деформированного текста. Сначала проводится сегментация слов путем нахождения пробелов между ними, используя метод Кэнни. Далее на изображении применяют три различных фильтра: RGB, Wavelet, Gradient. На следующем этапе начинаются два параллельных процесса: расчет матриц вероятности и увеличение фрагмента изображения со словом. Следующими итерациями процессов являются: расчет условной вероятности и расчет априорной вероятности. Получив данные от обоих процессов классификатор прогнозирует потенциальные границы каждого слова на изображении, после чего с помощью программы Tesseract происходит распознавание текста.

В процессе распознавания текста нередко складывается ситуация, когда система пытается распознать текст там, где его нет. Так, в статье [12] для решения проблемы смещения внимания распознавателя авторы предлагают метод – FAN (Focusing Attention Network), который задействует механизм фокусировки внимания на символах, чтобы автоматически предотвратить смещение внимания распознавателя. Система FAN состоит из двух компонентов: сеть внимания – AN (Attention network), которая отвечает за распознавание контрольных символов, и фокусирующая сеть – FN (Focusing network), которая отвечает за корректировку внимания путем оценки информации, которую AN вычленяет на определенных областях изображения.

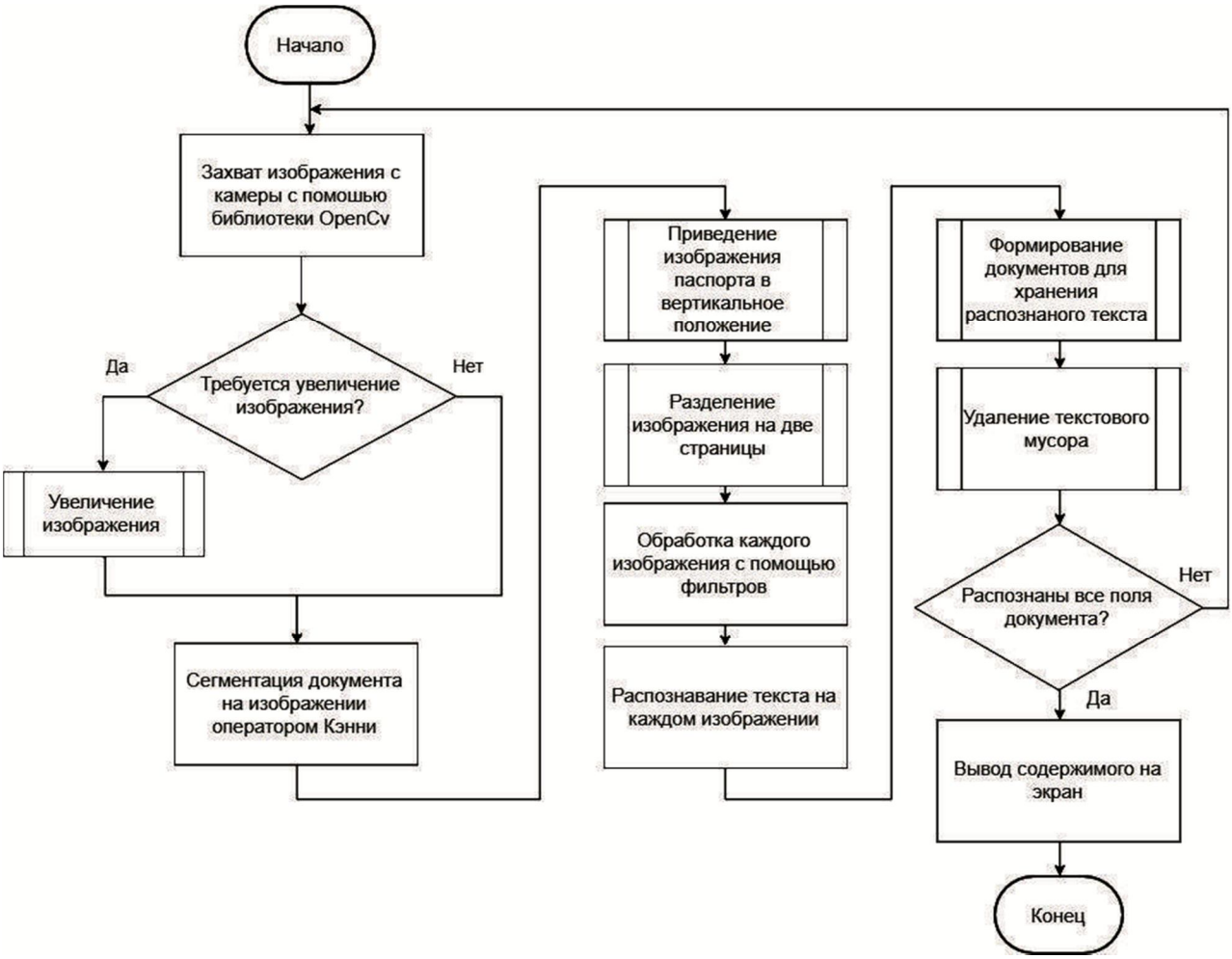
На российском рынке основной программой по распознаванию текста на документах является разработка Smart IDReader компании Smart Engines [13]. Данная программа является проприетарным программным обеспечением. В связи с этим не представляется возможным сравнение методов и алгоритмов, используемых данной программой, с разработанным нами методом.

Задача оптического распознавания символов остается актуальной в настоящее время. Проведенный выше анализ методов распознавания текста на изображении показал, что большинство разработанных алгоритмов направлены на работу с отсканированными документами, которые избавлены от бликов, появляющихся на изображениях, сделанных камерой. Благодаря наличию открытых инструментов для разработки аналогичной программы актуальным является доработка существующих программных средств для их адаптации под кириллический текст.

Разработка алгоритма для распознавания текста реквизитов паспорта РФ

Большинство существующих исследований, связанных с распознаванием текста, выполнены зарубежными компаниями и ориентированы на работу с языками, использующими латиницу. Это обусловлено тем, что латинский алфавит является «алфавитом международного общения», т.е. основой для письменности на многих языках. В силу изложенных фактов сохраняется актуальность разработок, связанных с распознаванием кириллицы, в частности, для работы с русским языком. Так, в рассмотренной выше статье [3] авторы предпринимают попытки разработать собственный инструмент для решения данной задачи. Мы решили прибегнуть к усовершенствованию готовых наработок, предназначенных для распознавания текста, а именно использовать программу с открытым исходным кодом Tesseract с применением дополнительных методов обработки изображения. На данный момент Tesseract представляет собой одно из лучших и легкодоступных программных средств, которое позволяет легко перенести текст с изображения в электронный документ. Как и многие другие средства, разрабатываемые для распознавания текста, Tesseract основан на рекуррентной нейронной сети – RNN (Recurrent neural network). Благодаря такому подходу, мы, в отличие от авторов [3], избегаем строгой привязки к определенным шрифтам, так как словари, поставляемые вместе с

Tesseract содержат очень большой объем данных по многим языкам, включая русский. На рисунке представлена блок-схема разработанного метода.



Алгоритм работы метода
Method Algorithm

Работа с веб-камерой осуществляется с помощью функций из библиотеки OpenCV: cv2.VideoCapture() для захвата изображения и cv2.imwrite() для сохранения его в заданной директории.

Далее по мере необходимости осуществляется увеличение изображения. Под увеличением понимается настройка камеры таким образом, чтобы документ занимал большую часть изображения.

После получения изображения происходит сегментация документа оператором Кэнни. Затем производится удаление посторонних шумов с изображения и «закрытие» границ пикселей – окрашивание в один и тот же цвет соседних пикселей, которые потенциально являются продолжением границы контура. Эта операция позволяет вычленить документ независимо от того, как он был расположен оператором.

Операция определения границ с помощью оператора Кэнни оказалась самым ресурсозатратным с вычислительной точки зрения процессом обработки изображения. Для повышения скорости работы программы было принято решение заменить данный шаг другим методом. Для этого было предложено повысить уровень цифрового увеличения изображения до такой степени, чтобы помимо паспорта, на изображении было как можно меньше посторонних артефактов (отражение подсветки на стеклянной поверхности, блики, составные части прибора).

Процедура приведения документа в вертикальное положение основана на результате, полученном на шаге сегментации. После определения границ паспорта инициализируются маски для самого документа и остального фона, где фон задается маской черного цвета. Затем высчитывается угол отклонения паспорта от горизонтальной оси и осуществляется поворот до требуемого положения с помощью функции rotate_bound() из библиотеки imutils.

В связи с тем, что разработанный метод ориентирован на работу с паспортами граждан РФ, которые содержат основную информацию о владельце паспорта на 2 отдельных страницах, была реализована функция разделения изображения документа на 2 страницы. Функция рассчитывает центр изображения и задает относительные координаты для каждой страницы.

Далее происходит процедура наложения фильтров на изображения с помощью функций из библиотеки skimage: cvtColor(), bilaterialFilter(), createCLAHE(), а также методов img_as_ubite() и mean_bilateral().

Для перевода текста из формата изображения была использована программа Tesseract со специальным словарем из открытого доступа и функцией image_to_string().

Для уменьшения вероятности ошибок при распознавании распознанный текст проходит обработку инструментом для проверки правописания jamspell [14]. Во многих контекстах этот инструмент удобен для коррекции ошибочно распознанных слов. Также текст проходит фильтрацию на предмет наличия так называемого «мусора» (ошибочно распознанных символов или слов/букв) с помощью функции jamspell.FixFragment().

Распознанный и синтаксически исправленный текст с каждой страницы сохраняется в отдельный текстовый файл. Помимо используемого ранее инструмента для корректировки распознанного текста осуществляется обработка символов, сохраненных в документе. В связи с тем, что при записи текста в файл могут возникнуть нежелательные символы или артефакты, осуществляется их удаление с помощью регулярных выражений и функций из библиотеки re.

После прохода всех шагов алгоритма осуществляется проверка на полноту распознанных данных. Под полнотой данных понимается успешное распознавание всего текста из ожидаемых полей документа (в данном случае: место выдачи паспорта, дата выдачи и т.д.). Если проверка пройдена успешно, то результаты работы программы классифицируются по категориям: ФИО, место выдачи, дата рождения, пол и т.д., это необходимо для обеспечения организованного вывода результата работы программы. Все распознанные данные выводятся на экран, в противном случае проводится повторное выполнение алгоритма.

Тестирование

Для проведения экспериментов по распознаванию изображений был разработан планшетный сканирующий стенд. Стенд представляет собой конструкцию закрытого бокса (коробки), в верхнюю поверхность которого установлено прозрачное стекло. Внутренние поверхности окрашены в черный матовый цвет для исключения паразитных отражений света извне. На некотором расстоянии от стекла по внутреннему периметру бокса расположена светодиодная лента подсветки с регулируемой яркостью. Цифровая USB-камера устанавливается на подвижный стол, с помощью которого регулируется расстояние до планшетного стекла. В проекте использовалась веб-камера Logitech Brio 4K Stream Edition, но возможна поддержка любой другой цифровой веб-камеры. Питание светодиодной подсветки осуществляется через разъем USB. Кабели цифровой камеры и питания подсветки выведены в нижней части стенда.

Для запуска программы необходимо подключить сканирующий стенд через USB кабель к компьютеру и запустить разработанную программу, которая примет снимок с веб-камеры, установленной в стенде, и выполнит распознавание текста.

Для проверки эффективности работы метода был проведен эксперимент, где расстояние между камерой и документом составило 45 см. В ходе тестирования при обнаружении границ паспорта возникли проблемы: на изображении из-за подсветки, используемой на стенде,

появились отражения и блики из-за глянцевого верхнего покрытия страниц паспорта. При заданном расстоянии между камерой и документом программа правильно распознала на страницах паспорта лишь малый процент текста.

Блики и отражение подсветки на изображении существенно ухудшали качество распознаваемого текста. В связи с этим были проведены эксперименты, в ходе которых на прозрачную поверхность, на которой лежит документ, накладывался антибликовый светофильтр. В результате удалось избавиться от бликов и отражений, но качество распознавания текста оставалось на низком уровне.

В следующем эксперименте при расстоянии между камерой и документом в 45 см применялось цифровое увеличение изображения до 25 см. Данный метод пагубно сказался на качестве снимка, что обеспечило низкое качество символов, что обусловило низкое качество символов на изображении.

Для решения проблем, возникших в ходе экспериментов, было предложено сократить расстояние между камерой и документом до 25 см. Такой подход позволил повысить качество снимка без применения светофильтра и без цифрового увеличения. В ходе проведения данного эксперимента разработанный метод распознавания текста обеспечил правильное распознавание почти всех полей документа, что свидетельствует об эффективности данного метода.

Далее представлена таблица с результатами экспериментов при использовании различных методов. Для определения эффективности работы каждого метода было проведено по 5 опытов. В качестве результата для каждого поля высчитывался средний показатель корректности распознавания текста в процентах.

Результаты экспериментов при различных методах распознавания текста
Experimental results obtained with various text recognition methods

	Средний показатель корректности распознавания текста, %		
Поле / Метод	Положение камеры на расстоянии 45 см без цифрового увеличения	Положение камеры на расстоянии 45 см с цифровым увеличением до расстояния в 25 см	Положение камеры на расстоянии 25 см от документа
Место выдачи	20	40	100
Дата выдачи	0	20	80
Код подразделения	0	0	80
Фамилия	0	20	80
Имя	20	40	60
Отчество	0	20	100
Пол	0	0	60
Дата рождения	0	20	100

В результате тестирования программы было выявлено, что самый высокий процент успешного распознавания полей паспорта достигается при расстоянии от камеры до документа в 25 см – процент правильно распознанной информации для каждого поля составил не меньше 60 %, тогда как при цифровом увеличении камеры и без него результаты составили не более 40 % и 20 % соответственно. Средний процент успешного распознавания всего текста на изображении при расстоянии между камерой и документом в 25 см составил 84,4 %, при других параметрах 24,4 % (цифровое увеличение до 25 см) и 8,8 % (расстояние 45 см).

Проведенные эксперименты позволили повысить скорость работы программы за счет сокращения ресурсозатратных процедур, связанных с обработкой изображения; с помощью физического приближения камеры к документу удалось уменьшить размеры сканирующего стенда и повысить качество распознавания текста.

Заключение

В данной работе представлены метод распознавания текста на изображении страниц паспорта гражданина РФ и перевод извлеченной информации в текстовый формат. Удалось достигнуть положительного результата, не прибегая к созданию и обучению собственных нейронных сетей. Для тестирования метода был разработан планшетный стенд с закреплённой на нем веб-камерой Logitech Brio 4K Stream Edition, при этом возможно крепление любой другой цифровой веб-камеры. Был проведен ряд экспериментов для выявления практических недостатков данного метода. По результатам тестирования было принято решение сделать акцент не на усовершенствовании имеющихся решений в области распознавания текста, а на манипуляциях непосредственно с изображением (его предобработкой), что дало положительный результат.

Для устранения возникших в ходе экспериментов проблем, таких как: низкое качество символов, блики, отражения от подсветки на глянцевой поверхности страниц паспорта, было предложено применить антибликовые светофильтры и сократить расстояние между камерой и документом до 25 см, без применения цифрового увеличения. В результате процент успешного распознавания текста составил 84,4 %, что в несколько раз выше, чем при использовании цифрового увеличения.

В дальнейшем предполагается продолжить модернизацию разработанного метода: улучшить используемые алгоритмы обработки изображений и распознавания текста, усовершенствовать аппаратную часть, сделав её более компактной, а также провести структурные и программные оптимизации кода для повышения скорости работы. В перспективе представляется возможным использовать стационарные модули, оборудованные данным программным обеспечением, на пропускных пунктах предприятий или в местах, где требуется ввод, идентификационных/паспортных данных.

Библиографический список

1. **Ye Q., Doermann D.** Text detection and recognition in imagery: A survey // IEEE transactions on pattern analysis and machine intelligence. 2015, vol. 37, no. 7, pp. 1480-1500.

2. **Breuel T. M.** High performance text recognition using a hybrid convolutional-lstm implementation // Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on. IEEE. 2017, vol. 1, pp. 11-16.

3. **Gunawan D., Arisandi D., Ginting F. M., Rahmat R. F., Amalia A.** Russian character recognition using Self-Organizing Map // Journics: Confce Series. IOP Publishing. 2017, vol. 801, no. 1, p. 012040.

4. **Uchida S.** Statistical Deformation Model for Handwritten Character Recognition // Advances in Digital Document Processing and Retrieval. 2014, pp. 157-174.

5. **Busta M., Neumann L., Matas J.** Deep textspotter: An end-to-end trainable scene text localization and recognition framework // Proceedings of the IEEE International Conference on Computer Vision. 2017, pp. 2204-2212.

6. **Марковников Н. М., Кипяткова И. С.** Аналитический обзор интегральных систем распознавания речи // Труды СПИИРАН. 2018. №. 58 (3). С. 77-110.

7. **Otsu N.** A threshold selection method from gray-level histograms // IEEE transactions on systems, man, and cybernetics. 1979, vol. 9, no. 1, pp. 62-66.

8. **Kohonen T.** The self-organizing map // Proceedings of the IEEE. 1990, vol. 78, no. 9, pp. 1464-1480.

9. **Shi B., Bai X., Yao C.** An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition // IEEE transactions on pattern analysis and machine intelligence. 2017, vol. 39, no. 11, pp. 2298-2304.

10. **Shi B., Wang X., Lyu P., Yao C., Bai X.** Robust scene text recognition with automatic rectification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, pp. 4168-4176.

11. **Roy S., Shivakumara P., Roy P. P., Pal U., Tan C. L., Lu T.** Bayesian classifier for multi-oriented video text recognition system // Expert Systems with Applications. 2015, vol. 42, no. 13, pp. 5554-5566.

12. **Cheng Z., Cheng Z., Bai F., Xu Y., Zheng G., Pu S., Zhou S.** Focusing attention: Towards accurate text recognition in natural images // Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE. 2017, pp. 5086-5094.

13. Smart Engines [Электронный ресурс]. URL: https://smartengines.ru/about/, (дата обращения: 10.08.2019).

14. JamSpell [Электронный ресурс]. URL: https://github.com/bakwc/JamSpell, (дата обращения: 15.08.2019).

UDC 004.724

A SPECIALIZED METHOD OF TEXT RECOGNITION FOR AUTOMATIC PROCESSING OF PASSPORT DATA

P. S. Drugov, Junior researcher of SPIIRAS, St. Petersburg, Russia; Master's student of ETU, St. Petersburg, Russia;
orcid.org/0000-0002-0319-0554, e-mail: pah53k@gmail.com
E. E. Usina, Junior researcher of SPIIRAS, St. Petersburg, Russia; Master's student of SUAI, St. Petersburg, Russia;
orcid.org/0000-0001-9745-0216, e-mail: lizzzi96@mail.ru

Smart environments, being developed for further implementation within existing enterprise architectures, require a crucial prerequisite to be met: adding of identity data of new visitors into enterprise database. ***The aim of*** *this paper is to develop a method for text recognition on the pages of internal passport of the Russian Federation. The aim of the developed method presented here consists in the processing of page content (on the example of Russian internal passport) including, particularly, the following operations: image magnification, object (passport page) segmentation, image processing with filters and text recognition. To test the method a vertical stand with space for placing a document on top was designed accompanied by a relevant source code in Python programming language. The tests performed within this research revealed, that the highest success rate in text recognition is achieved when the distance from a camera to a document is 250 mm (88,8 %); whereas by camera focus adjustment and distance increase the recognition outcomes were significantly worse, 66,7 % and 22,2 % respectively. Implementing such solution into enterprise operation would allow avoiding human mistake factor in data transfer (caused by poor attention or fatigue of employees) and to optimize the input of customer data into database.*

Key words: *text recognition, OCR, Tesseract, optical character recognition, OpenCV, image processing, domestic passport, smart environment.*

DOI: 10.21667/1995-4565-2019-70-43-51

References

1. **Ye Q., Doermann D.** Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence*. 2015, vol. 37, no. 7, pp. 1480-1500.

2. **Breuel T. M.** High performance text recognition using a hybrid convolutional-lstm implementation. Document Analysis and Recognition (ICDAR). *14th IAPR International Conference on. IEEE*. 2017, vol. 1, pp. 11-16.

3. **Gunawan D., Arisandi D., Ginting F. M., Rahmat R. F., Amalia A.** Russian character recognition using Self-Organizing Map. *Journics: Confce Series. IOP Publishing*. 2017, vol. 801, no. 1, p. 012040.

4. **Uchida S.** Statistical Deformation Model for Handwritten Character Recognition // Advances in Digital Document Processing and Retrieval. 2014, pp. 157-174.

5. **Busta M., Neumann L., Matas J.** Deep textspotter: An end-to-end trainable scene text localization and recognition framework. Proceedings of the *IEEE International Conference on Computer Vision*. 2017, pp. 2204-2212.

6. **Markovnikov N. M., Kipyatkova I. S.** Analiticheskij obzor integral'nyh sistem raspoznavaniya rechi. (An analytic survey of end-to-end speech recognition systems), *Trudy SPIIRAN*, 2018, 3(58), pp. 77-110. (in Russian).

7. **Otsu N.** A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*. 1979, vol. 9, no. 1, pp. 62-66.

8. **Kohonen T.** The self-organizing map. *Proceedings of the IEEE*. 1990, vol. 78, no. 9, pp. 1464-1480.

9. **Shi B., Bai X., Yao C.** An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE transactions on pattern analysis and machine intelligence*. 2017, vol. 39, no. 11, pp. 2298-2304.

10. **Shi B., Wang X., Lyu P., Yao C., Bai X.** Robust scene text recognition with automatic rectification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 4168-4176.

11. **Roy S., Shivakumara P., Roy P. P., Pal U., Tan C. L., Lu T.** Bayesian classifier for multi-oriented video text recognition system. *Expert Systems with Applications*. 2015, vol. 42, no. 13, pp. 5554-5566.

12. **Cheng Z., Cheng Z., Bai F., Xu Y., Zheng G., Pu S., Zhou S.** Focusing attention: Towards accurate text recognition in natural images. *Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE*. 2017, pp. 5086-5094.

13. Smart Engines [elektronnyj resurs]. URL: https://smartengines.ru/about/, (data of access: 10.08.2019).

14. JamSpell [elektronnyj resurs]. URL: https://github.com/bakwc/JamSpell, (data of access: 15.08.2019).