# Метод сегментации структурированных текстовых объектов на изображении с помощью динамического программирования*

М. А. Поволоцкий[I,II,III], Д. В. Тропин[II,III], Т. С. Чернов[III], Б. И. Савельев[III]

[I] Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, г. Москва, Россия
[II] Московский физико-технический институт (государственный университет), г. Долгопрудный, Россия
[III] Смарт Энджинс Сервис, г. Москва, Россия

**Аннотация.** Рассматривается задача сегментации изображений текстовых фрагментов с известными ограничениями на взаимное расположение элементов. Рассматривается модель, в которой граф ограничений является простой цепью. Показано, что задача сегментации в этом случае может быть решена точно с помощью алгоритма динамического программирования, причем этот алгоритм обладает оптимальной асимптотической трудоемкостью. Данный алгоритм был встроен в две системы распознавания. Первая система предназначена для распознавания документов, удостоверяющих личность, таких как паспорта и водительские права. Вторая система предназначена для распознавания автомобильных номеров, в ней предложенный алгоритм был использован для сегментации номеров на отдельные символы. Проведены эксперименты на закрытом наборе данных по замеру качества и производительности полученных решений на мобильном телефоне. Экспериментальные результаты показали, что полученные решения превосходят по качеству алгоритмы, не использующие ограничения на взаимное расположение элементов, а их трудоемкость позволяет работать на мобильных устройствах в режиме реального времени.

**Ключевые слова:** сегментация текста, динамическое программирование, распознавание документов, обработка изображений, OCR.

**DOI** 10.14357/20718632190306

## Введение

В современном мире возникает все больше задач, требующих автоматизированной оцифровки документов, чье количество и разнообразие неизменно растет. Для удовлетворения потребностей индустрии от исследователей требуется разработка новых методов распознавания документов [1], в частности, позволяющих работу в режиме реального времени непосредственно на мобильных устройствах [2].

Неотъемлемым этапом решения задачи оцифровки документа является сегментация текста – разбиение всех пикселей изображения на области, соответствующие текстовым фрагментам и фону. Выбор подхода для решения задачи сегментации изображения текста зависит от большого числа факторов: условий регистрации, гибкости геометрической структуры документа, доступных вычислительных ресурсов и т.д. В данной работе объектами сегмента-

*Работа выполнена при финансовой поддержке РФФИ, гранты №17-29-07092 и №17-29-03236

ции выступают документы, удостоверяющие личность, такие как паспорта и водительские права, а также автомобильные номера; условия съемки предполагаются неконтролируемыми; в качестве устройства для съемки и распознавания используется мобильное устройство (смартфон, планшет) без доступа к сторонним вычислительным ресурсам.

Про документы, удостоверяющие личность, известно, из скольких текстовых строк и полей они состоят и в каких пределах могут варьироваться размеры полей и расстояния между ними. А для автомобильных номеров многих стран, в том числе Российской Федерации, известно точное расположение символов в зависимости от принадлежности номера определенному типу из заданного набора. Особенностью задачи распознавания номеров является малое число вертикальных и горизонтальных линий на изображении, что делает сложной задачу, предшествующую сегментации, – проективное исправление области интереса. Следовательно, от алгоритма сегментации номеров требуется устойчивость к проективным искажениям, которые можно моделировать изменениями горизонтальных и вертикальных смещений между символами в некоторых пределах. Вопрос расчёта возможных смещений при известной неточности проективного исправления рассматривается, например, в работе [3].

В случае работы на мобильном устройстве в режиме реального времени алгоритм сегментации текста должен быть устойчив к вариациям освещенности, возможным в неконтролируемых условиях съемки, и иметь высокое быстродействие.

Существующие методы сегментации текста отличаются друг от друга необходимым объемом априорной информации и, как следствие, универсальностью. Так, многие методы включают в себя предварительную бинаризацию изображения текста, например, в целях последующего анализа компонент связности и формирования из них отдельных фрагментов текста [4]. В основе этих методов лежит предположение, что пиксели текста отделимы от пикселей фона порогом по яркости, глобальным или локальным. В рассматриваемых задачах данное предположение часто оказывается


Рис. 1. Примеры работы известных алгоритмов бинаризации

неверным из-за неравномерного освещения, неоднородной окраски фона, наличия посторонних объектов на изображении и т.д. (Рис. 1).

Другой тип априорного знания о сегментируемом тексте – представление о его внутренней геометрической структуре. Например, в задаче распознавания отсканированных форм и бланков с высокой точностью известно положение каждого текстового фрагмента. Задача сегментации в этом случае становится тривиальной [5]. Однако в нашей задаче данный подход неприменим из-за большей вариативности расположения сегментируемого текста.

Для сегментации гибких структурированных объектов используют шаблонное описание, в котором для каждого элемента структуры задана модель, позволяющая оценивать правдоподобность расположения сегмента в некоторой области изображения, а также заданы ожидаемые расстояния между элементами и штрафы за деформацию расстояний относительно ожидаемых. Такой подход позволяет сегментировать объекты самой разной природы. Например, в работе [6] выполняется выделение на изображениях людей, животных, автомобилей. В работе [7] решается задача прослеживания лиц на видео с помощью покадрового выделения лица, модель которого задана шаблоном. В работе [8] на основе известной геометрии контура производится выделение ядра клетки. Текстовые объекты тоже сегментируются с помощью шаблонного описания: например, в [9] используется трехстрочный шаблон для выделения текстовых полей при распознавании банковских карт.

Существуют методы, предназначенные для сегментации текстов произвольной структуры, основанные на анализе вертикальных и горизонтальных проекций [10]. Такие методы предполагают только то, что текстовые строки горизонтальны, а структура колонок известна. Еще более слабым допущением является гипотеза о том, что горизонтальные границы текстовых строк лежат на параллельных прямых. Методы, базирующиеся на этой гипотезе, заключаются в поиске особенностей на образе изображения текста в пространстве Хафа [11]. Наконец, существуют алгоритмы «text in the wild», не использующие никакой априорной информации. Они могут быть реализованы как с помощью искусственных нейронных сетей [12], так и классических методов обработки изображений [13], а также их комбинации [14]. Данные алгоритмы обладают большой универсальностью, но имеют низкое быстродействие.

Поскольку документы, удостоверяющие личность, и автомобильные номера обладают геометрической структурой с известными ограничениями на смещения между информационными полями и символами соответственно, нами был выбран подход, основанный на шаблонном описании.

## 1. Сегментация составных объектов с ограничениями на взаимное расположение элементов с помощью динамического программирования

В работе [15] описано, как можно формально поставить задачу сегментации составного объекта с известными ограничениями на взаимное расположение его элементов. Нас интересует случай, когда граф ограничений является простой цепью.

Перечислим входные данные алгоритма. Для каждого возможного положения *i*-го элемента $l_i$ из конечного множества мощности $W$ задается значение функции штрафа $m_i(l_i)$ тем меньшее, чем выше правдоподобие того, что *i*-й элемент находится в положении $l_i$. Кроме того, для каждой комбинации положений $l_i$ и $l_{i+1}$ соседних *i*-го и *i*+1-го элементов задается значение функции штрафа за деформацию связи между ними $d_i(l_i, l_{i+1})$. Число элементов равно $N$. Размер входных данных составляет $O(NW^2)$ за счет функций $d_i(l_i, l_{i+1})$.

Выходом алгоритма является набор положений всех элементов $l = l_1, \ldots, l_N$, такой, что минимально значение следующего функционала:

$$Z(l) \equiv \sum_{i=1}^{N} m_i(l_i) + \sum_{i=1}^{N-1} d_i(l_i, l_{i+1}) \to \min_{l_1,\ldots,l_N}. \quad (1)$$

Такую задачу можно решить алгоритмом динамического программирования, предлагаемым в [15]. Данный алгоритм имеет трудоемкость $O(NW^2)$.

Докажем, что задачу (1) можно решить алгоритмом меньшей трудоемкости, а именно $O(NW)$, если положения $l_i$, $i = 1, \ldots, N$ описываются одной координатой, а функции $d_i(l_i, l_{i+1})$ имеют следующий вид:

$$d_i(l_i, l_{i+1}) = \begin{cases} 0, & T_{min}(i) \leqslant l_{i+1} - l_i \leqslant T_{max}(i), \\ +\infty, & \text{иначе;} \end{cases} \quad (2)$$

где $T_{min}(i)$ и $T_{max}(i)$ — предельно допустимые значения смещения между положениями $i$-го и $i+1$-го элементов. В этом случае, добавив в задачу ограничения (2), можно упростить функционал (1) следующим образом:

$$Z(l) = \sum_{i=1}^{N} m_i(l_i) \to \min_{l},$$
$$T_{min}(i) \leqslant l_{i+1} - l_i \leqslant T_{max}(i), \quad i = 1, \ldots, N-1. \quad (3)$$

Необходимо обратить внимание на то, что, во-первых, введенная модель деформации (2) исключает возможность искажений, выходящих за пределы $T_{min}$ и $T_{max}$, а во-вторых, любые допустимые искажения не штрафуются.

Фактический размер входных данных в результате понизился до $O(NW)$, поскольку каждая функция штрафа за деформацию $d_i(l_i, l_{i+1})$ задается не набором значений на всем множестве возможных аргументов размером $W^2$, а парой чисел $T_{min}(i)$ и $T_{max}(i)$, и наибольший объем занимают значения функций $m_i(l_i)$. Значит, выполняется необходимое

условие существования алгоритма с трудоемкостью $O(NW)$.

В процессе решения задачи (3) с помощью динамического программирования заполняются две таблицы: таблица $L(i,j)$ оптимальных положений $(i-1)$-го элемента при условии, что $i$-ый элемент находится в положении $j$, и таблица $C(i,j)$ кумулятивных стоимостей первых $i$ элементов с тем же условием. Проинициализируем первую строку таблицы $C$ значениями функции $m_1(j)$:

$$C(1,j) = m_1(j), \qquad j = 1,\ldots,W. \tag{4}$$

Далее схема заполнения таблиц следующая:

$$\begin{aligned} L(i+1,j) &= \underset{j-T_{max}(i)\leqslant j'\leqslant j-T_{min}(i)}{\arg\min} C(i,j'), \; i = 1,\ldots N-1, \\ C(i+1,j) &= C(i,L(i+1,j)) + m_{i+1}(j), \; j = 1,\ldots,W. \end{aligned} \tag{5}$$

После заполнения таблиц оптимальное положение $l_N$ последнего элемента определяется как индекс глобального минимума последней строки таблицы $C$:

$$l_N = \underset{1\leqslant j\leqslant W}{\arg\min} C(N,j). \tag{6}$$

Оптимальные положения остальных элементов $l_1,\ldots,l_{N-1}$ вычисляются обратным проходом по таблице $L$:

$$l_i = L(i+1,l_{i+1}), \quad i = N-1,\ldots,1. \tag{7}$$

Минимум функционала $Z(l)$ определяется значением глобального минимума в последней строке таблицы $C$:

$$Z_{min} \equiv \min_l Z(l) = C(N,l_N). \tag{8}$$

Оценим трудоемкость алгоритма. Заполнение $i$-й строки таблиц $L$ и $C$ требует $W$-кратного вычисления минимума на отрезке одной и той же длины $T_{max}(i) - T_{min}(i) + 1$. Данную задачу можно решить за время $O(W)$ с помощью различных методов, например, используя алгоритм ван Херка/Гила-Вермана [16]. Тогда полное заполнений таблицы потребует $O(NW)$ операций. Поиск глобального максимума последней строки таблицы $C$ требует $O(W)$ операций, обратный проход по таблице $L$ — $O(N)$ операций. Итого, общая трудоемкость алгоритма составляет $O(NW)$, что и требовалось доказать. Заметим, что асимптотическая трудоемкость алгоритма оптимальна, так как пропорциональна размеру входных данных.

## 2. Алгоритм выделения информационных полей документа

Для многих документов, удостоверяющих личность, характерно наличие текстовой зоны большой площади, содержащей базовую информацию: полное имя, дату рождения, пол и т.д., далее по тексту будем называть ее зоной основного заполнения (Рис. 2). Её структура схожа для документов одного типа: совпадающие число строк, схожие межстрочные расстояния, одинаковое число текстовых фрагментов, соответствующих информационным полям, и, наконец, примерно равные расстояния между ними. Обычно эта информация заранее известна. Построим формальную модель шаблона документа, в котором эта априорная информация описана.

Шаблон документа задает упорядоченную сверху вниз последовательность $\{\phi_i\}_{i=1}^N$ горизонтальных полос двух типов: «строка» и «пропуск», чередующихся между собой. Длина последовательности $N$ известна. Для каждой полосы $\phi_i$ заданы минимальная и максимальная высоты $h_i^{min}$ и $h_i^{max}$. Кроме того, каждая полоса $\phi_i$ типа «строка» обладает внутренней сруктурой, заданной в виде последовательно-

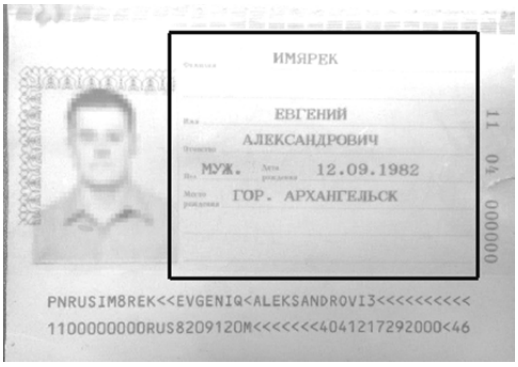



Рис. 2. Страница паспорта РФ с зоной основного заполнения, выделенной прямоугольником

сти $\{\psi_{i,j}\}_{j=1}^{M_i}$ чередующихся блоков, также двух типов: «поле» и «пропуск». Длины последовательностей $M_i$ тоже известны. Для каждого блока $\psi_{i,j}$ заданы предельные значения ширины $w_{i,j}^{min}$ и $w_{i,j}^{max}$. Первая горизонтальная полоса в шаблоне, а так же первый блок в каждой «строке» являются «пропусками». «Полосы» и «блоки» не пересекаются и покрывают изображение полностью. Далее будем опускать кавычки в названиях типов полос и блоков.

Таким образом, для того, чтобы некоторый набор $L = \left\{t_i, b_i, \{l_{i,j}, r_{i,j}\}_{j=1}^{M_i}\right\}_{i=1}^{N}$ верхних и нижних границ $[t_i, b_i)$ для каждой полосы $\phi_i$, а также левых и правых границ $[l_{i,j}, r_{i,j})$ для каждого блока $\psi_{i,j}$ удовлетворял шаблону, должны выполняться следующие условия:

- высоты полос и ширины блоков лежат внутри заданных в шаблоне диапазонов

$$h_i^{min} \leqslant b_i - t_i \leqslant h_i^{max}, \quad i = 1,\ldots,N,$$
$$w_{i,j}^{min} \leqslant r_{i,j} - l_{i,j} \leqslant w_{i,j}^{max},\ j = 1,\ldots,M_i,\ i = 1,\ldots,N; \tag{9}$$

- между полосами и блоками отсутствуют промежутки

$$b_i = t_{i+1}, \quad i = 1,\ldots,N-1,$$
$$r_{i,j} = l_{i,j+1}, \quad j = 1,\ldots,M_i - 1, \quad i = 1,\ldots,N; \tag{10}$$

- границы краевых полос и блоков совпадают с границами изображения:

$$t_1 = 0, \quad b_N = H,$$
$$l_{i,1} = 0, \quad r_{i,M_i} = W, \qquad i = 1,\ldots,N, \tag{11}$$

где $W$ и $H$ — ширина и высота изображения, соответственно. Ограничения (10) и (11) позволяют задавать положение каждой полосы одним числом вместо двух, например, только $t_i \equiv y_i$ для полос и $l_{i,j} \equiv x_{i,j}$ для блоков. Остальные границы определяются тривиально. Модель шаблона построена.

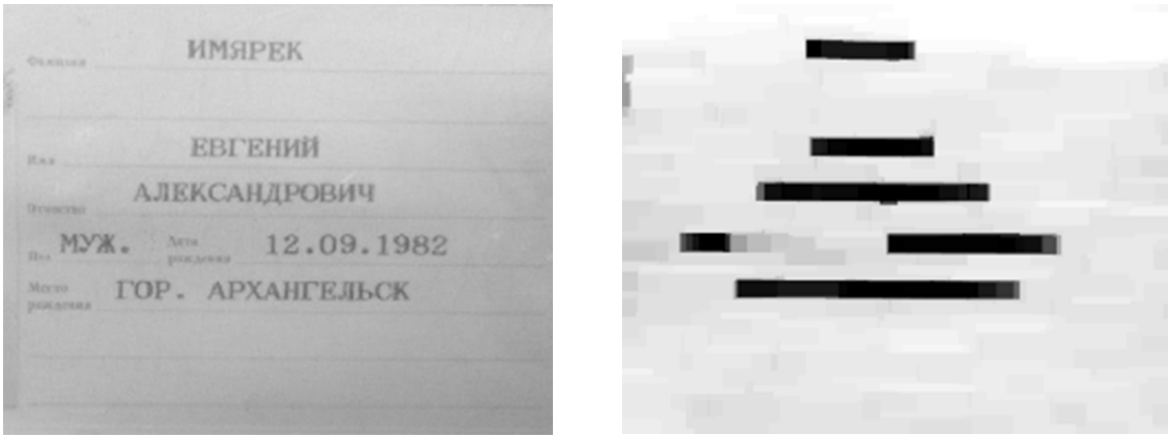
Входными данными задачи выделения полей являются нормализованное ахроматическое изображение зоны основного заполнения $I(x,y)$ (Рис. 3 а) и вышеописанный шаблон. Ответом задачи является набор границ $L = \left\{t_i, b_i, \{l_{i,j}, r_{i,j}\}_{j=1}^{M_i}\right\}_{i=1}^{N}$ такой, что информационные поля лежат внутри границ блоков и положения границ удовлетворяют шаблону.

Сформулируем задачу как оптимизационную. Выбор оптимизационного критерия основан на предположении, что средняя яркость текста на изображении документа меньше средней яркости фона. Чем больше текста и чем меньше фона попало внутрь полей, тем лучше результат их выделения, поэтому для оптимизационной задачи был выбран такой функционал, который оптимален при высокой средней яркости пикселей вне полей и низкой — внутри, а именно межклассовая дисперсия [17], домноженная на знак разности средних яркостей классов:

$$V = \omega_0\ \omega_1(\mu_0 - \mu_1)^2 \operatorname{sgn}(\mu_0 - \mu_1) = \\ = \omega_0\ \omega_1(\mu_0 - \mu_1)\,|\,\mu_0 - \mu_1\,| \to \max, \tag{12}$$

где $\omega_0$ и $\omega_1$ — доли нулевого и первого классов от всех пикселей изображения, $\mu_0$ и $\mu_1$ — средние внутриклассовые значения яркости. К первому классу относятся пиксели внутри полей, к нулевому — пиксели внутри пропусков, как горизонтальных, так и вертикальных.

На начальном этапе алгоритма выделения полей предварительно обработаем входное изображение так, чтобы фоновые пиксели внутри полей сравнялись по яркости с пикселями текста, а яркость пикселей внутри пропусков осталась существенно отличной. Целью такой предобработки является увеличение разности средних яркостей классов, полученных корректным разбиением, по отношению к уровню шума на изображении. Воспользовавшись информацией из шаблона, определим следующие величины: $h_{max}$ – наибольшая высота текста, $h_{min}$ – наименьшая высота текста, $w_{min}$ – наименьшая ширина пропуска. Последовательно применим к изображению следующие операции: морфологическое замыкание с квадратным структурным элементом размера $\left(2\left\lceil\frac{h_{max}}{2}\right\rceil+1\right)\times\left(2\left\lceil\frac{h_{max}}{2}\right\rceil+1\right)$, вычитание входного изображения из результата замыкания, морфологическое размыкание инвертированной разности с горизонтальным структурным




(a) (b)

Рис. 3. Входное изображение (a) и результат предварительной обработки (b)

элементом $\left(2\left\lceil\frac{w_{min}}{2}\right\rceil-1\right)\times 1$, морфологическое замыкание предыдущего изображения с вертикальным структурным элементом $1\times\left(2\left\lfloor\frac{h_{min}}{2}\right\rfloor-1\right)$ и автоконтрастирование. Пример результата предобработки приведен на Рис. 3 b.

Далее будем оптимизировать функционал (12), вычисляемый по предобработанному изображению, в два этапа. На первом будем считать размеры строк и полей фиксированными. При этом задача максимизации величины V становится эквивалентной минимизации суммарной яркости пикселей внутри полей. Чтобы это доказать, перепишем (12) в терминах суммарных яркостей классов $S_0$ и $S_1$ и объемов $Q_0$ и $Q_1$:

$$\omega_0=\frac{Q_0}{Q_0+Q_1},\quad \omega_1=\frac{Q_1}{Q_0+Q_1},\quad \mu_0=\frac{S_0}{Q_0},\quad \mu_1=\frac{S_1}{Q_1},$$

$$V=\frac{Q_0Q_1}{(Q_0+Q_1)^2}\left(\frac{S_0}{Q_0}-\frac{S_1}{Q_1}\right)\left|\frac{S_0}{Q_0}-\frac{S_1}{Q_1}\right|\to\max. \tag{13}$$

Очевидно, что величина $S=S_0+S_1$ постоянна и равна сумме яркостей всех пикселей изображения $I(x,y)$. Выразим через неё $S_0$ в формуле (13):

$$V=\frac{Q_0Q_1}{(Q_0+Q_1)^2}\left(\frac{S-S_1}{Q_0}-\frac{S_1}{Q_1}\right)\left|\frac{S-S_1}{Q_0}-\frac{S_1}{Q_1}\right|=$$

$$=\frac{Q_1}{Q_0(Q_0+Q_1)^2}\left(S-\frac{Q_0+Q_1}{Q_1}S_1\right)\left|S-\frac{Q_0+Q_1}{Q_1}S_1\right|\to\max \tag{14}$$

При фиксированных размерах полей $Q_0$ и $Q_1$ не меняются, поэтому $V$ является функцией одной переменной $S_1$. $V$ монотонно убывает с ростом $S_1$, следовательно, минимизация $S_1$ приводит к максимизации $V$, что и требовалось доказать.

Таким образом, мы пришли к следующей оптимизационной задаче:

$$S_1=\sum_{k=1}^{\lfloor N/2\rfloor}\sum_{m=1}^{\lfloor M_{2k}/2\rfloor}I_\Sigma(x_{2k,2m},y_{2k},w_{2k,2m},h_{2k})\to\min, \tag{15}$$

где $I_\Sigma(x,y,w,h)$ — суммарная яркость в прямоугольном окне с левым верхним углом в точке $(x,y)$, шириной $w$ и высотой $h$. Заметим, что её значение может быть вычислено для любых допустимых значений аргументов за константное число операций, если предварительно вычислить интегральное изображение [18] за время $O(WH)$. В сумме используются прямоугольники только с чётными индексами, поскольку горизонтальные полосы и вертикальные блоки с нечётными индексами являются пропусками.

Предположим, что для любого положения $y$ $k$-й строки известна минимальная суммарная яркость внутри её полей

$$S_k(y) = \min_x \sum_{m=1}^{\lfloor M_{2k}/2 \rfloor} I_\Sigma(x_{2k,2m}, y, w_{2k,2m}, h_{2k}). \quad (16)$$

Тогда задача (15) сводится к

$$\sum_{k=1}^{\lfloor N/2 \rfloor} S_k(y_{2k}) \to \min_y. \quad (17)$$

С учетом (9) на значения $y$ накладываются ограничения

$$h_{2k} + h_{2k+1}^{min} \leqslant y_{2k+2} - y_{2k} \leqslant h_{2k} + h_{2k+1}^{max},$$
$$k = 1, \ldots, \left\lfloor \frac{N-1}{2} \right\rfloor. \quad (18)$$

Таким образом, задача приобрела вид (3), что позволяет применить наш алгоритм для её решения. Ограничения на положения первой и последней «строк» можно задать бесконечным штрафом:

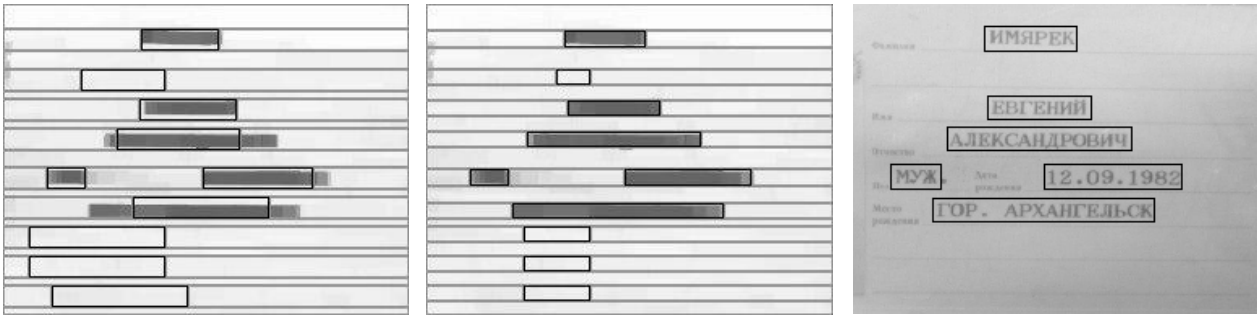
$$S_2'(y) = \begin{cases} S_2(y), & h_1^{min} \leqslant y \leqslant h_1^{max}, \\ +\infty, & \text{иначе;} \end{cases}$$
$$S_{N-1}'(y) = \begin{cases} S_{N-1}(y), & H - h_N^{max} \leqslant y + h_{N-1} \leqslant H - h_N^{min}, \\ +\infty, & \text{иначе.} \end{cases} \quad (19)$$

Задача (16) вычисления оптимального выделения полей внутри одной строки в фиксированном положении сводится к задаче (3) аналогичным образом. Значит, задачу (15) выделения строк и полей фиксированных размеров можно решить динамическим программированием в два этапа: найти оптимальное положение полей для каждого возможного положения каждой строки, а затем — оптимальное положение самих строк (Рис. 4 a).

На втором этапе для уточнения размеров и положений полей предлагается максимизировать (12) покоординатным спуском: поочередно перебирая поля и их границы в рамках шаблона, увеличивая целевой функционал (Рис. 4 b). В качестве начального приближения воспользуемся решением, полученным на первом этапе. Общая асимптотическая трудоемкость алгоритма составляет $O(WH\sum_{k=1}^{\lfloor N/2 \rfloor} M_{2k})$ (вычисление интегрального изображения занимает $O(WH)$, динамическое программирование внутри строк — $O(WH\sum_{k=1}^{\lfloor N/2 \rfloor} M_{2k})$, динамическое программирование по строкам — $O(HN)$, покоординатный спуск — $O((W+H)\sum_{k=1}^{\lfloor N/2 \rfloor} M_{2k})$).

Финальные оценки положений полей изображены прямоугольниками на Рис. 4 c.

Построенный алгоритм сегментации был внедрен в индустриальную систему распознавания документов Smart IDReader [19]. Для проверки применимости алгоритма в реальных условиях были проведены эксперименты по оценке качества распознавания выделенных полей. В работе [20] тестирование сегментации полей документа проводилось на закрытом наборе из 7700 изображений российских паспортов с использованием точности распознавания выделенных полей для оценки качества сегментации. Предложенный алгоритм сравнивался с алгоритмом на основе проекций, описанным в работе [10] и использующим инфор-




(a) (b) (c)

Рис. 4. Поэтапные результаты сегментации зоны основного заполнения для документов, удостоверяющих личность

(a) максимизация (12) при фиксированном размере полей
(b) максимизация (12) покоординатным спуском
(c) результат сегментации

Табл. 1. Качество распознания сегментированных текстовых полей документов

| Метод | Имя | Фамилия | Дата рождения | Дата выдачи | Место рождения |
|---|---|---|---|---|---|
| Внутренние российские паспорта | | | | | |
| Слугин и др. [9] | 0,0646 | 0,0715 | 0,0768 | 0,0686 | 0,0767 |
| Предложенный | **0,0356** | **0,0451** | **0,0533** | **0,0543** | **0,0495** |
| Водительские права России | | | | | |
| Слугин и др. | 0,0987 | 0,0885 | 0,0956 | 0,1384 | 0,3003 |
| Предложенный | **0,0519** | **0,0563** | **0,0712** | **0,0585** | **0,1445** |
| Водительские права Великобритании | | | | | |
| Слугин и др. | 0,0753 | 0,0774 | 0,0837 | 0,0979 | 0,1266 |
| Предложенный | **0,0243** | **0,0231** | **0,0749** | **0,0269** | **0,0738** |

мацию не о структуре документов, а только о предельных размерах, одинаковых для всех строк и полей. В этой работе размер набора изображений паспортов РФ был расширен до 10410 экземпляров и были добавлены 2591 и 713 изображений водительских прав России и Великобритании, соответственно; общий объем данных составил, таким образом, 13714 изображений документов. Для оценки качества распознавания одного поля использовалась нормализованная метрика Левенштейна [21]:

$$V(r,w)=\frac{2\cdot \text{levenshtein}(r,w)}{|r|+|w|+\text{levenshtein}(r,w)}, \quad (24)$$

где $r$ — результат распознавания текстового поля, $w$ — корректное текстовое поле, взятое из разметки, $\text{levenshtein}(r,w)$ — расстояние Левенштейна между двумя строками. Чем меньше значение метрики, тем выше качество распознавания. Метрика была заменена для того, чтобы с большим весом учитывать ошибки, с большей вероятностью сделанные на этапе сегментации, а не на последующем этапе распознавания уже выделенных полей.

Оценка качества работы обоих алгоритмов проводилась независимо для документов разного типа, на подмножестве полей, общих для всех рассматриваемых типов. В Табл. 1 приведены значения оценок качества распознавания, усредненные по всем экземплярам одного поля. Из таблицы видно, что использование нашего алгоритма выделения полей приводит к росту качества распознавания по сравнению с базовым алгоритмом на всех полях. Примеры выделения полей на разных документах, а так же зоны основного заполнения проиллюстрированы Рис 5.

Также были проведены замеры времени работы системы Smart IDReader в целом и выполнения этапа сегментации в частности. Было установлено, что общее время работы всей системы, запущенной на iPhone 5s в однопоточном режиме, составило 5 минут 43 секунды (на наборе из 100 изображений), из которых на выделение полей было затрачено 7,1 секунды. Таким образом, этап сегментации занял приемлемые 2%.

## 3. Сегментация автомобильных номеров

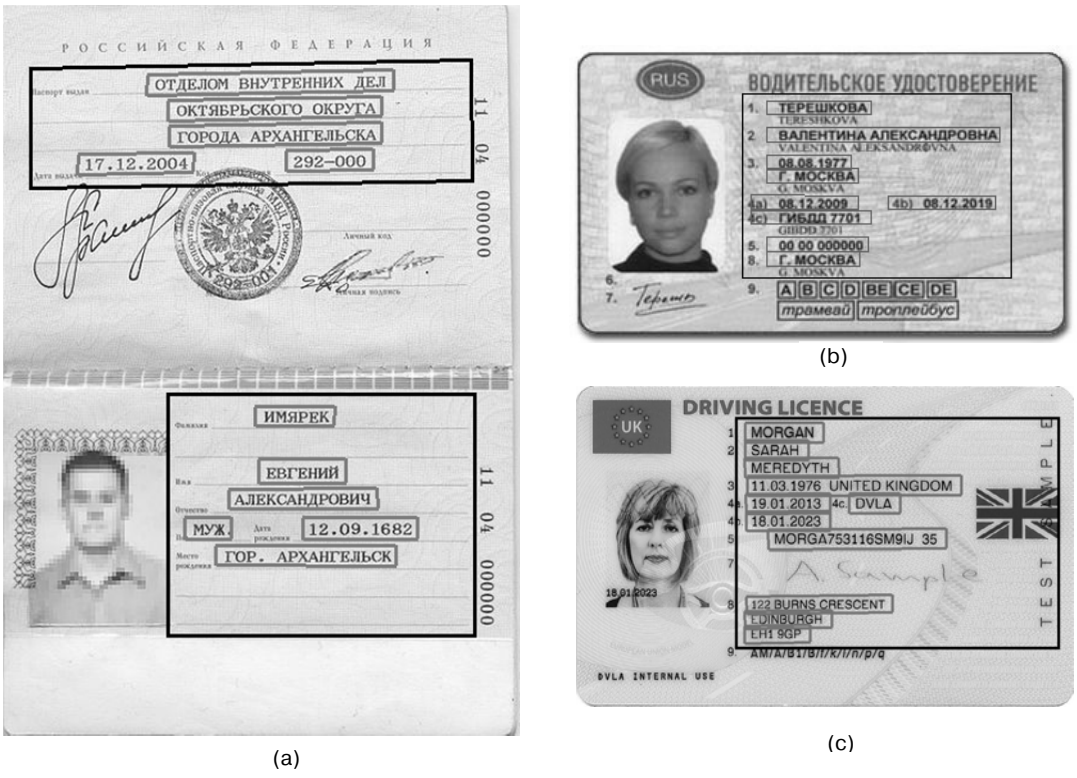
Во многих странах формат автомобильных номеров подчиняется определенному стандарту: существует один или более типов, каждый из которых обладает фиксированным расположением символов. Предполагая, что тип номера известен, построим модель шаблона номера.

Зададим в шаблоне начальные положения прямоугольников, содержащих символы, – знакомест $p_i^0=(x_i^0,y_i^0,w_i^0,h_i^0)$ и наложим следующие ограничения на их изменения:

- знакоместа не выходят за пределы изображения:

$$0\leqslant x_i\leqslant W-w_i,\ 0\leqslant y_i\leqslant H-h_i,\ i=1,\ldots,N; \quad (21)$$

- предельное изменение расстояний между соседними знакоместами пропорционально евклидову расстоянию между центрами их начальных положений:




(a) (b) (c)

Рис. 5. Пример зон основного заполнения и полей
(a) для внутреннего паспорта РФ
(b) для новых водительских прав РФ
(c) для водительских прав Великобритании

$$
\begin{aligned}
&| x_{i+1} - x_i | \leqslant \delta \,|| \, c_{i+1}^0 - c_i^0 \, ||_2, \\
&| y_{i+1} - y_i | \leqslant \delta \,|| \, c_{i+1}^0 - c_i^0 \, ||_2, \qquad (22) \\
&c_i^0 = \left( x_i^0 + \frac{w_i^0}{2}, y_i^0 + \frac{h_i^0}{2} \right), \; i = 1, \ldots, N;
\end{aligned}
$$

- размеры знакомест не меняются:

$$w_i = w_i^0, \qquad h_i = h_i^0, \qquad i = 1, \ldots, N. \qquad (23)$$

В качестве функционала, к оптимизации которого мы сводим задачу сегментации, по аналогии с задачей поиска строк и полей документа используется суммарная яркость пикселей изображения номера $I(x, y)$ внутри знакомест $p_i = (x_i, y_i, w_i, h_i)$:

$$\sum_{i=1}^{N} \sum_{x,y \in p_i} I(x, y) = \sum_{i=1}^{N} I_\Sigma(p_i) \to \min, \qquad (20)$$

где $N$ — количество символов. Считается, что средняя яркость символов меньше средней яркости фона; если для рассматриваемого типа номеров это неверно, изображение предварительно инвертируется. Морфологическая фильтрация входного изображения в качестве предобработки в данной задаче неприменима, поскольку размеры участков фона внутри знакомест и в промежутках между ними имеют один порядок. Для увеличения контраста между пикселями вне и внутри знакомест воспользуемся автоконтрастированием.

Получившаяся задача (20) сводится к виду (3), если зафиксировать одну из координат у всех знакомест, что позволяет поочередно оптимизировать $x_i$ и $y_i$ нашим алгоритмом. Итеративный подбор положений знакомест проиллюстрирован Рис. 6. Более подробно алгоритм сегментации описан в работах [22, 23].

Асимптотическая трудоемкость получившегося алгоритма составляет $O(WH)$, причем наиболее вычислительно сложным оказывается вычисление интегрального изображения (оптимизация по $x$ занимает $O(NW)$, по $y$ — $O(NH)$, число итераций невелико).

Предложенный алгоритм сегментации был встроен в систему распознавания автомобильных

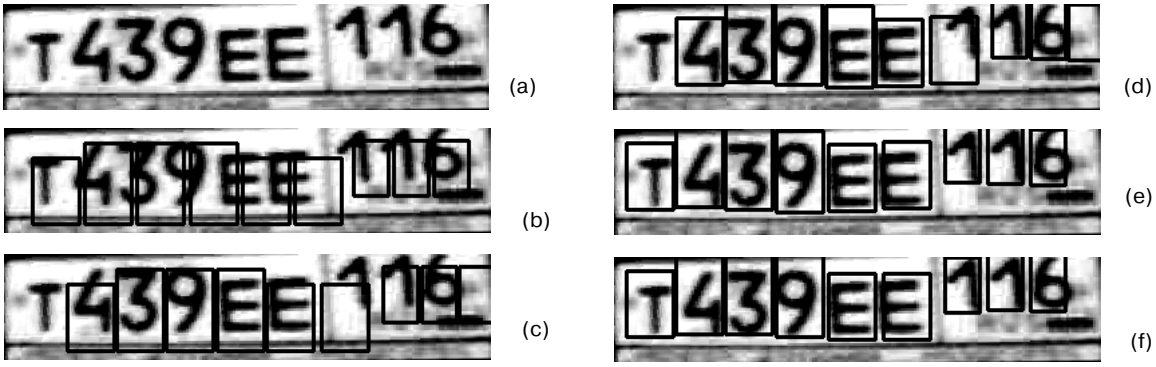



Рис. 6. Итеративная подстройка положений знакомест (под одной итерацией подразумевается подстройка по одной координате)

(a) входное изображение номера
(b) начальное расположение знакомест
(c)-(f) итерация 1-4

Табл. 2. Качество распознавания автомобильных номеров в зависимости от жесткости ограничений (лучший результат выделен жирным)

| Значение $\delta$ | 0.0 | 0.02 | 0.05 | 0.1 | 0.5 | 2.0 |
|---|---|---|---|---|---|---|
| Доля ошибок | 0.0553 | 0.0505 | **0.0491** | 0.0545 | 0.1027 | 0.1090 |

номеров МАРИНА [24]. Тестирование сегментации номеров проводилось на наборе из 7591 изображений автомобильных номеров РФ. Тип каждого номера считался заранее известным. Была исследована зависимость качества распознавания номера после сегментации от величины параметра δ, задающего ограничения на смещение соседних знакомест; нулевому значению δ соответствует полный запрет на смещения, большим значениям δ – отсутствие ограничений. Нарушение порядка между символами и наложение символов друг на друга не допускалось при любых значениях δ. В качестве метрики оценки качества использовалась доля ошибочно распознанных символов от общего числа символов в тестовом наборе. Результаты экспериментов представлены в Табл. 2. Как видно из таблицы, использование ограничений позволяет добиться уменьшения числа ошибок на 11% по сравнению с запретом на относительные смещения символов и более чем в два раза по сравнению с допущением произвольного расположения символов.

Было измерено время, затраченное на сегментацию, а также общее время работы системы МАРИНА, запущенной на iPhone 5s в однопоточном режиме на наборе из 100 изображений. Замеры показали, что этап сегментации занял 2 секунды из суммарных 27, т.е. 7.4%.

## Заключение

В работе предложен метод сегментации изображений текстовых фрагментов с известными ограничениями на взаимное расположение элементов с помощью динамического программирования. Алгоритм был применен для решения задач выделения информационных полей документов, удостоверяющего личность, и сегментации автомобильных номеров. Результаты экспериментов показали преимущество предложенного подхода в сравнении с методами, не использующими априорную информацию об ограничениях. Возможным направлением дальнейших исследований является модификация алгоритма в целях работы с объектами более сложной структуры. Кроме того, интересен вопрос применимости алгорита для решения оптимизационной задачи с другими функциями штрафов.

## Литература


1. Nagy G. Disruptive developments in document recognition //Pattern Recognition Letters. – 2016. – V. 79. – P. 106-112.
2. Арлазаров, В. В., Жуковский, А. Е., Кривцов, В. Е., Николаев, Д. П., Полевой, Д. В. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов //Информационные технологии и вычислительные системы. – 2014. – Т. 3. – С. 71-81.
3. Konovalenko I. A., Shemiakina J. A. Error values analysis for inaccurate projective transformation of a quadrangle //Journal of Physics: Conference Series. – IOP Publishing, 2018. – V. 1096. – №. 1. – P. 012038.
4. Feldbach M., Tönnies K. D. Robust Line Detection in Historical Church Registers //Joint Pattern Recognition Symposium. – Springer, Berlin, Heidelberg, 2001. – P. 140-147.
5. Арлазаров В. В., Постников В. В., Шоломов Д. Л. Cognitive Forms-система массового ввода структурированных документов //Труды Института системного анализа Российской академии наук. – 2002. – Т. 1. – С. 35-46.
6. Felzenszwalb, P. F., Girshick, R. B., McAllester, D., Ramanan, D. Object detection with discriminatively trained part-based models //IEEE transactions on pattern analysis and machine intelligence. – 2010. – V. 32. – №. 9. – P. 1627-1645.
7. Chrysos, G. G., Antonakos, E., Zafeiriou, S., Snape, P. Offline deformable face tracking in arbitrary videos //Proceedings of the IEEE International Conference on Computer Vision Workshops. – 2015. – P. 1-9.
8. Zhang, L., Kong, H., Liu, S., Wang, T., Chen, S., Sonka, M. Graph-based segmentation of abnormal nuclei in cervical cytology //Computerized Medical Imaging and Graphics. – 2017. – V. 56. – P. 38-48.
9. Sheshkus, A., Nikolaev, D. P., Ingacheva, A., Skoryukina, N. Approach to recognition of flexible form for credit card expiration date recognition as example //Eighth International Conference on Machine Vision (ICMV 2015). – International Society for Optics and Photonics. – 2015. – V. 9875. – P. 98750R.
10. Слугин Д. Г., Арлазаров В. В. Поиск текстовых полей документа с помощью методов обработки изображений //Труды Института системного анализа Российской академии наук. – 2017. – Т. 67. – №. 4. – С. 65-73.
11. Louloudis, G., Gatos, B., Pratikakis, I., Halatsis, C. Text line and word segmentation of handwritten documents //Pattern Recognition. – 2009. – V. 42. – №. 12. – P. 3169-3183.
12. Wang K., Belongie S. Word spotting in the wild //European Conference on Computer Vision. – Springer, Berlin, Heidelberg, 2010. – P. 591-604.
13. Epshtein B., Ofek E., Wexler Y. Detecting text in natural scenes with stroke width transform //2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – IEEE, 2010. – P. 2963-2970.
14. Turki H., Halima M. B., Alimi A. M. A hybrid method of natural scene text detection using MSERs masks in HSV space color //Ninth International Conference on Machine Vision (ICMV 2016). – International Society for Optics and Photonics, 2017. – V. 10341. – P. 1034111.
15. Felzenszwalb P. F., Zabih R. Dynamic programming and graph algorithms in computer vision //IEEE transactions on pattern analysis and machine intelligence. – 2011. – V. 33. – №. 4. – P. 721-740.
16. Van Herk M. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels //Pattern Recognition Letters. – 1992. – V. 13. – №. 7. – P. 517-521.
17. Otsu N. A threshold selection method from gray-level histograms //IEEE transactions on systems, man, and cybernetics. – 1979. – V. 9. – №. 1. – P. 62-66.
18. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features //Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. – IEEE, 2001. – V. 1. – P. I-I.
19. Bulatov K. B., Arlazarov V. V., Chernov T. S., Slavin O. A., Nikolaev D. P. Smart IDReader: Document Recognition in Video Stream // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – IEEE, 2017. – V. 6. – P. 39-44. ISSN 2379-2140, ISBN 978-15-38635-86-5, doi: 10.1109/ICDAR.2017.347.
20. Povolotskiy M. A., Tropin D. V. Dynamic programming approach to template-based OCR //Eleventh International Conference on Machine Vision (ICMV 2018). – International Society for Optics and Photonics, 2019. – V. 11041. – P. 110411T.
21. Yujian L., Bo L. A normalized Levenshtein distance metric //IEEE transactions on pattern analysis and machine intelligence. – 2007. – V. 29. – №. 6. – P. 1091-1095.
22. Povolotskiy M. A., Kuznetsova E. G., Khanipov T. M. Russian license plate segmentation based on dynamic time warping //European Conference on Modelling and Simulation. – 2017. – P. 285-291.
23. Поволоцкий М. А., Кузнецова Е. Г., Уткин Н. В., Николаев Д. П.. Сегментация регистрационных номеров автомобилей с применением алгоритма динамической трансформации временной оси //Сенсорные системы. – 2018. – Т. 32. – № 1. – С. 50-59. doi: 10.7868/S0235009218010080
24. Visillect. МАРИНА — Модуль распознавания номеров автомобилей. 2019. URL: http://visillect.com/ru/alpr


**Поволоцкий Михаил Александрович**. Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), Москва, Россия. Младший научный сотрудник, магистр. Количество печатных работ: 8. Область научных интересов: обработка изображений, дискретная оптимизация. E-mail: mikhail.povolotskiy@iitp.ru

**Тропин Даниил Вячеславович**. Федеральное государственное автономное образовательное учреждение высшего профессионального образования "Московский физико-технический институт (государственный университет)" (МФТИ ГУ), Долгопрудный, Россия. Студент второго курса магистратуры. Бакалавр. Количество печатных работ: 4. Область научных интересов: машинное зрение, анализ и обработка изображений, распознавание документов. E-mail: tropin.dv@phystech.edu

**Чернов Тимофей Сергеевич**. Федеральное государственное учреждение "Федеральный исследовательский центр "Информатика и управление" Российской академии наук (ФИЦ ИУ РАН), Москва, Россия. Младший научный сотрудник. Кандидат технических наук. Количество печатных работ: 30. Область научных интересов: системное программирование, компьютерное зрение, распознавание документов, оценка качества изображений. E-mail: chernov.tim@smartengines.com

**Савельев Борис Игоревич**. ООО "Смарт Энджинс Сервис", Москва, Россия. Научный сотрудник – программист. Магистр. Количество печатных работ: 6. Область научных интересов: искусственный интеллект, машинное обучение, системы распознавания, информационные технологии. E-mail: bsaveliev@smartengines.com

## Dynamic Programming Approach to Textual Structured Objects Segmentation in Images

M. A. Povolotskiy [I, II, III], D. V. Tropin [II, III], T. S. Chernov [III], B. I. Savelyev [III]

[I] The Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia
[II] Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia
[III] Smart Engines Service, Moscow, Russia

**Abstract.** This paper deals with the problem of segmentation of images of text fragments with known constraints on the relative position of elements. The model in which the constraints form a path graph is considered. It is shown that the segmentation problem in this case can be solved precisely with use of a dynamic programming algorithm, and this algorithm has an optimal asymptotic complexity. This algorithm was built into two recognition systems. The first system was designed to recognize identity documents, such as passports and driver's licenses. The proposed algorithm was used in this system to extract information fields. To do this, a two-level field hierarchy was introduced, in which the fields were grouped in rows, within which they were ordered from left to right, and the lines themselves were ordered from top to bottom. The second system was designed to recognize license plates in which the proposed algorithm was used to segment plates into individual characters. In this case, the natural ordering of characters from left to right was introduced. Thus, the generality of the proposed approach is demonstrated. Experiments were conducted on a closed data set to measure the quality and performance of the solutions obtained on a mobile phone. Experimental results showed that the solutions obtained are superior in quality to algorithms that do not use constraints on the mutual arrangement of elements, and their complexity allows them to work on mobile devices in real time.
**Keywords:** text segmentation, dynamic programming, document recognition, image processing, OCR.

**DOI** 10.14357/20718632190306

## References

1. Nagy G. Disruptive developments in document recognition //Pattern Recognition Letters. – 2016. – V. 79. – P. 106-112.
2. Arlazarov V.V., Zhukovsky A.E., Krivtsov V.E., Nikolaev D.P., Polevoy D.V. Analiz osobennostey ispol'zovaniya statsionarnykh i mobil'nykh malorazmernykh tsifrovykh video kamer dlya raspoznavaniya dokumentov [Analysis of features of the use of fixed and mobile small-sized digital video camera for OCR]. //Informatsionnye tekhnologii i vychislitel'nye sistemy [Journal of Information Technologies and Computing Systems] – 2014. – V. 3. – P. 71-81.
3. Konovalenko I. A., Shemiakina J. A. Error values analysis for inaccurate projective transformation of a quadrangle //Journal of Physics: Conference Series. – IOP Publishing, 2018. – T. 1096. – №. 1. – C. 012038.
4. Feldbach M., Tönnies K. D. Robust Line Detection in Historical Church Registers //Joint Pattern Recognition Symposium. – Springer, Berlin, Heidelberg, 2001. – P. 140-147.
5. Arlazarov V. V., Postnikov V. V., Sholomov D.L. Cognitive Forms – sistema massovogo vvoda strukturirovannykh dokumentov [Cognitive Forms – system for mass input of structured documents] //Trudy Instituta sistemnogo analiza rossiyskoy akademii nauk [Proceeding of the Institute for Systems Analysis of the Russian Academy of Science]. – 2002. – V. 1. – P. 35-46.
6. Felzenszwalb, P. F., Girshick, R. B., McAllester, D., Ramanan, D. Object detection with discriminatively trained part-based models //IEEE transactions on pattern analysis and machine intelligence. – 2010. – V. 32. – №. 9. – P. 1627-1645.

7. Chrysos, G. G., Antonakos, E., Zafeiriou, S., Snape, P. Offline deformable face tracking in arbitrary videos //Proceedings of the IEEE International Conference on Computer Vision Workshops. – 2015. – P. 1-9.
8. Zhang, L., Kong, H., Liu, S., Wang, T., Chen, S., Sonka, M. Graph-based segmentation of abnormal nuclei in cervical cytology //Computerized Medical Imaging and Graphics. – 2017. – V. 56. – P. 38-48.
9. Sheshkus, A., Nikolaev, D. P., Ingacheva, A., Skoryukina, N. Approach to recognition of flexible form for credit card expiration date recognition as example //Eighth International Conference on Machine Vision (ICMV 2015). – International Society for Optics and Photonics. – 2015. – V. 9875. – P. 98750R.
10. Slugin D. G., Arlazarov V. V. Poisk tekstovykh poley dokumenta s pomoshch'yu metodov obrabotki izobrazheniy [Text fields extraction based on image processing] //Trudy Instituta sistemnogo analiza rossiyskoy akademii nauk [Proceeding of the Institute for Systems Analysis of the Russian Academy of Science]. – 2017. – V. 67. – № 4. – P. 65–73.
11. Louloudis, G., Gatos, B., Pratikakis, I., Halatsis, C. Text line and word segmentation of handwritten documents //Pattern Recognition. – 2009. – V. 42. – №. 12. – P. 3169-3183.
12. Wang K., Belongie S. Word spotting in the wild //European Conference on Computer Vision. – Springer, Berlin, Heidelberg, 2010. – P. 591-604.
13. Epshtein B., Ofek E., Wexler Y. Detecting text in natural scenes with stroke width transform //2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – IEEE, 2010. – P. 2963-2970.
14. Turki H., Halima M. B., Alimi A. M. A hybrid method of natural scene text detection using MSERs masks in HSV space color //Ninth International Conference on Machine Vision (ICMV 2016). – International Society for Optics and Photonics, 2017. – V. 10341. – P. 1034111.
15. Felzenszwalb P. F., Zabih R. Dynamic programming and graph algorithms in computer vision //IEEE transactions on pattern analysis and machine intelligence. – 2011. – V. 33. – №. 4. – P. 721-740.
16. Van Herk M. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels //Pattern Recognition Letters. – 1992. – V. 13. – №. 7. – P. 517-521.
17. Otsu N. A threshold selection method from gray-level histograms //IEEE transactions on systems, man, and cybernetics. – 1979. – V. 9. – №. 1. – P. 62-66.
18. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features //Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. – IEEE, 2001. – V. 1. – P. I-I.
19. Bulatov K. B., Arlazarov V. V., Chernov T. S., Slavin O. A., Nikolaev D. P. Smart IDReader: Document Recognition in Video Stream // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – IEEE, 2017. – V. 6. – P. 39-44., ISSN 2379-2140, ISBN 978-15-38635-86-5, doi: 10.1109/ICDAR.2017.347.
20. Povolotskiy M. A., Tropin D. V. Dynamic programming approach to template-based OCR //Eleventh International Conference on Machine Vision (ICMV 2018). – International Society for Optics and Photonics, 2019. – V. 11041. – P. 110411T.
21. Yujian L., Bo L. A normalized Levenshtein distance metric //IEEE transactions on pattern analysis and machine intelligence. – 2007. – V. 29. – №. 6. – P. 1091-1095.
22. Povolotskiy M. A., Kuznetsova E. G., Khanipov T. M. Russian license plate segmentation based on dynamic time warping //European Conference on Modelling and Simulation. – 2017. – P. 285-291.
23. Povolotskiy M.A., Kuznetsova E.G., Utkin N.V, Nikolaev D.P. Segmentatsiya registratsionnykh nomerov avtomobiley c primeneniem algoritma dinamicheskoy transformatsii vremennoy osi [Segmentation of vehicle registration plates based on dynamic time warping] //Sensornye sistemy [Sensory systems]. – 2018. – V. 32. – № 1. – P. 50–59. doi: 10.7868/S0235009218010080
24. Visillect. MARINA — Automatic license plate recognition system. Available at: http://visillect.com/en/alpr/ (accessed June 17, 2019)

**Povolotskiy M. A.** Master. The Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoy Karetny per. 19, build.1, Moscow, 127051, Russia, e-mail: mikhail.povolotskiy@iitp.ru

**Tropin D. V.** Bachelor. Moscow Institute of Physics and Technology (State University), Institutskiy Pereulok, 9, Dolgoprudny, Moskovskaya oblast', 141701, Russia, e-mail: tropin.dv@phystech.edu

**Chernov T. S.** Ph.D. Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia, e-mail: chernov.tim@smartengines.com

**Savelyev B. I.** Master. LLC "Smart Engines Service", Moscow, Russia, e-mail: bsaveliev@smartengines.com