

Распознавание документов, удостоверяющих личность, с помощью автономных мобильных устройств

Т.И. Булдакова¹, О.А. Славин²

¹Московский физико-технический институт (государственный университет)

²Институт системного анализа РАН

Введение

В последние несколько лет большое внимание уделяется исследованию вопросов распознавания различных объектов, в частности текстовых документов, с использованием мобильных телефонов, как в качестве устройств захвата изображения, так и в качестве вычислительных устройств. Так в разных работах обычно подходят к рассмотрению каких-то подзадач, например, таких, как оценка качества изображений, отдельные задачи распознавания, постобработка полученных результатов. При этом серьезного рассмотрения процесса в целом, а особенно возникающих обратных связей уделяется значительно меньше внимания. В основном рассматриваются системы, в которых захват кадра производится мобильным телефоном, а распознавание выполняется на сервере. В данной работе мы рассматриваем схему построения системы распознавания на устройстве, возникающие обратные связи и их применение для повышения качества распознавания.

Обзор проблемы и классическая схема распознавания документа

Рассматривается задача распознавания документов на примере паспорта гражданина РФ с помощью мобильного устройства. Исходными данными служит видеопоток (набор кадров), полученный камерой устройства. В отличие от сканера, документ может располагаться в некоторой произвольной плоскости относительно плоскости сфокусированного изображения, что добавляет задачу проективного преобразования полученных кадров. Задача распознавания документа, как правило, разделяется на следующие этапы:

- детектирование документа в кадре
- нахождение зоны и параметров проективного восстановления
- проективное исправление изображения документа
- позиционирование зон реквизитов (полей) на исправленном изображении
- распознавание реквизитов
- постобработка

В случае видеопотока задача постобработки является особенно актуальной. Межкадровая информация позволяет повысить точность и скорость получения результата распознавания. Мы остановимся на решении задачи интеграции результатов распознавания отдельных полей и на организации управления механизмом распознавания с целью оптимизации быстрейшего действия.

Схема предлагаемого подхода

Пусть для каждого отдельно взятого документа есть набор из n распознанных кадров I_0, I_1, \dots, I_{n-1} соответственно с m распознанными полями $F_0(I_i), F_1(I_i), \dots, F_{m-1}(I_i)$, а результат распознавания каждого поля $F_j(I_i)$ представлен набором знакомест

$$\{ \mathcal{N}_1(F_j(I_i)), \dots, \mathcal{N}_l(F_j(I_i)) \},$$

где l равняется либо 0 (соответствует случаю, когда в образе I_i поле $F_j(I_i)$ не найдено), либо количеству знакомест поля. Каждое знакоместо $\mathcal{N}_q(F_j(I_i))$ состоит из последовательности альтернатив

$$A_{q_1}(F_j(I_i)), \dots, A_{q_{1y}}(F_j(I_i)).$$

Число альтернатив u является количеством символов алфавита распознавания. Каждая альтернатива $A_{qu}(I_i)$ содержит код символа $s_{qu}(I_i)$ и его оценку $w_{qu}(I_i)$ в диапазоне $0 \div 1$.

Для простоты можно рассматривать одно единственное поле, тогда последовательность распознанных кадров, будет представлена в виде набора знакомест этого поля, найденных в каждом из кадров:

$$\{ \mathcal{N}_1(I_i), \dots, \mathcal{N}_l(I_i) \},$$

где l равняется либо 0 (соответствует случаю, когда в образе I_i поле не найдено), либо количеству знакомест поля, а каждое знакоместо $\mathcal{N}_q(I_i)$ содержит набор альтернатив

$$A_{q_1}(I_i), \dots, A_{q_{1y}}(I_i).$$

Рассмотрим представление распознанного поля в виде набора знакомест $A(I_i) = \{A_1(I_i), \dots, A_l(I_i)\}$, где $l=l(I_i)$ – количество распознанных знакомест в поле (это количество зависит от кадра), а каждое знакоместо $A_q(I_i)$ содержит код символа $s_q(I_i)$ и его оценку $w_q(I_i)$ в диапазоне $0 \div 1$. Два набора альтернатив $A(I_i)$ и $A'(I_i)$ считаются одинаковыми, если

$$s_q(I_i) = s'_q(I_i) \quad \forall q.$$

Значения $1-w_q(I_i)$ мы будем интерпретировать как вероятности ошибочного распознавания знакоместа $A_q(I_i)$.

Выделим подпоследовательность кадров I_{z_1}, \dots, I_{z_2} , $0 \leq z_1 < z_2 \leq n-1$, где z_1 – первый из кадров, в котором поле было найдено, z_2 – последний из кадров, в котором поле было найдено.

Для начала рассмотрим поля, состоящие из одного слова, например, фамилия, имя и другие. Для каждого $k \in [z_1, z_2]$ для поля, представленного в виде строки

$$S(I_i) = \{s_1(I_i), \dots, s_l(I_i)\}$$

и набора оценок $\{w_1(I_i), \dots, w_l(I_i)\}$ назначим оценку слову,

$$\Theta(S(I_i)) = \min_{1 \leq q \leq l} w_q(I_i),$$

которая необходима для кластеризации.

Будем рассматривать популяцию слов с оценками: $\{(S(I_{z_1}), \Theta(S(I_{z_1}))), \dots, (S(I_{z_2}), \Theta(S(I_{z_2})))\}$ и решать задачу выбора наилучшего результата по нескольким первым словам $S(I_{z_1}), \dots, S(I_r)$, минимизируя количества слов r . Необходимо отметить, что описанные выше характерные дефекты оцифровки не позволяют считать, что все строки $S(I_i)$ относятся к одному и тому прообразу. Прежде всего, возможны различные варианты определения границ отдельных символов и границ всего поля, поэтому в популяции строки $S(I_i)$ могут различаться.

Для решения задачу выбора наилучшего результата мы объединим одинаковые строки в кластеры. Каждый из кластеров Cl , относящихся к одной строке $S(Cl)$, состоит из количества строк, попавших в кластер, $|Cl|$ и оценку кластера $\Theta(Cl)$.

Вначале (то есть при первом обнаружении поля) создадим кластер Cl_1 из строки $S(I_{z_1})$. При дальнейшем обнаружении поля в кадре I_k (где $k > z_1$) строки, такой что $S(I_k) = S(I_x)$ (строки сравниваются посимвольно, $x \in [z_1, k]$) проводится модификация одного из существующих кластеров со строкой $S(I_x)$, либо создается новый кластер.

Для кластера, состоящего из одного элемента, оценка

$$\Theta(Cl) = \min_{1 \leq q \leq l} w_q(I_i) \quad (1)$$

является минимальной из оценок $w_q(I_i)$ альтернатив $A(I_i)$. При добавлении в кластер нового элемента с знакоместами $A'(I_i) = \{A'_1(I_i), \dots, A'_l(I_i)\}$ оценки рассчитываются следующим образом:

$$\Theta(Cl) = 1 - (1 - \Theta(Cl)) \cdot (1 - w'_q(I_i)), \quad (2)$$

а оценка нового кластера $\Theta(Cl_t)$ определяется по формуле (1).

На каждом шаге $k \in [z_1, z_2]$ рассмотрим кластеры Cl_1, \dots, Cl_p , состоящие из одинаковых наборов знакомест $A(I_i)$ и оценим их объемы $|Cl|$ и оценки $\Theta(Cl)$.

После очередного обновления кластеров на шаге t ($t > z_1$) проводится оценка состояния набора кластеров, состоящая в сравнения характеристик каждого из кластеров Cl с заранее заданными порогами:

$$|Cl| > \lambda_1, \quad \Theta(Cl) > \lambda_2. \quad (3)$$

При нахождении хотя бы одного кластера, для которого справедливо выполнение условий (3), процесс интеграции результатов распознавания завершается и в качестве результата интеграции берется $S(Cl)$. В случае, когда нашлось несколько кластеров, для которых справедливо выполнение условий (3), в качестве результата возьмем строку $S(Cl)$ из кластера с наибольшей оценкой $\Theta(Cl)$.

Для полей, состоящих из нескольких слов, например, поля "место рождения", предлагается обобщение алгоритма, включающее разбиение результата распознавания поля на отдельные слова и последующую конкатенацию выбранных "наилучших" вариантов слов.

Приемлемость предложенного алгоритма была проверена на тестовом наборе T_1 с хорошим и средним качеством видеосъемки паспортов РФ. В наборе T_1 содержались как видеопоследовательности с изображением паспорта гражданина РФ с четкими образами символов в кадрах, так и кадры со специфическими дефектами оцифровки на мобильных устройствах (смазывания, дефокусировка, блики из-за голографических элементов). Однако в T_1 не было видеопоследовательностей с подавляющим количеством кадров с дефектами оцифровки, т.е. человек может в каждой из видеопоследовательностей найти для каждого поля изображения, которые читаются глазом без затруднений. Набор состоял из 67 видеопоследовательностей объемом от 17 до 100 кадров. Анализировалось распознавание семи полей (ФИО, номер и серия паспорта, дата и место рождения), поле «место рождения» состояло из нескольких слов, остальные поля – из одного слова.

Для оценки предложенного алгоритма интеграции использовались две характеристики:

- *точность*, определяемая как доля результатов распознавания, совпадающих с истинным значением поля,
- *ускорение*, определяемое как среднее значение количества слов r , необходимых для интеграции, отнесенного к количеству кадров $(z_2 - z_1)$ видеопоследовательности,
- *среднее число кадров*, необходимое для распознавания каждого поля.

Характеристики алгоритма, полученные на наборе T_1 , сведены в таблицу 1, в которую также помещены комментарии по неудачной интеграции результатов.

Заключение

Отметим, что в основном ошибки алгоритма интеграции объясняются случаями отсутствия правильных результатов распознавания, а также ошибками поиска границ поля. Также отметим, что, несмотря на низкие средние значения ускорения реальные значения $r/(z_2 - z_1)$ для некоторых видеопоследовательностей точность близка к 1.

Что касается среднего числа кадров, необходимого для распознавания, то для поля «место рождения» получаем, что в среднем достаточно 8-9 кадров, для остальных полей – 3-5. Поле «место рождения», как правило, состоит из нескольких слов, и для каждого из них происходит выбор наилучшего. Поэтому полученное для него значение объясняется тем, что при отборе каждого слова число кадров, необходимое для распознавания и достижения при этом заданного уровня точности, может быть различным, и, следовательно, нужно взять наибольшее.

Таким образом, предложенный алгоритм интеграции позволяет получить высокую точность распознавания полей при существенном ограничении качества исходных данных для распознавания.

Таблица 1

Поле	Точность	Ускорение	Среднее число кадров	Комментарии по ошибкам
Фамилия	0,99	0,14	5,47	1 ошибка распознавания
Имя	0,99	0,1	4,59	1 ошибка распознавания
Отчество	0,99	0,11	4,91	1 ошибка распознавания
Серия паспорта	0,97	0,08	3,78	2 ошибки распознавания
Номер паспорта	0,99	0,08	3,68	1 ошибка распознавания
Дата рождения	0,96	0,08	4,34	2 ошибки распознавания
Место рождения	0,91	0,22	8,87	4 ошибки распознавания, 2 ошибки наведения рамки поля

Литература

1. 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. №3, 2014. С. 71-78.

2. *Арлазаров В.Л., Марченко А.А, Шоломов Д.Л.* Накопительные контексты в задаче распознавания// Труды ИСА РАН, 2014. Т. 64, №4. С. 64-72.
3. *I.S. Bangdiwala, R.J. Monroe* Some sequential procedures for ordering populations according to means, variances and regression coefficients //Institute of statistics Mimeo Series №202, 1958. P. 78.
4. *Shanti S. Gupta, S. Panchapakesan* On Sequential ranking and selection procedures. Technical Report #88-32C //Department of Statistic Purdue University, 1988. P. 30.
5. *Jayanti Chotai* Selection and rankong procedures based on likelihood rations //Department of Mathematical Statistics University of Umeå, 1979. P. 29