

Обработка и анализ изображений и сигналов

Определение типа сканированного документа методом динамической трансформации временных осей

Т.Р. МАКСИМОВА^I, П.В. БЕЗМАТЕРНЫХ^{I,II}

^I ООО «Смарт Энджинс Сервис», г. Москва, Россия

^{II} Федеральный исследовательский центр «Информатика и управление»
Российской академии наук, г. Москва, Россия

Аннотация. В работе решается задача выбора типа сканированного документа из предопределенного набора, возникающая в области автоматического распознавания изображений документов. Предлагаемый метод сопоставляет параллельные проекции входного изображения с эталонными проекциями образцов из предопределенного набора, для создания которых достаточно всего нескольких примеров. Сопоставление производится алгоритмом динамического выравнивания временных осей. Метод не требует бинаризации, распознавания ключевых слов или поиска геометрических примитивов, но необходима предварительная компенсации наклона документа. Проведенные эксперименты на изображениях деловых документов (8 типов документов) нормализованных вручную и автоматически показали точность типизации 99,79% и 99,76% соответственно. Среднее время обработки на ПК (AMD Ryzen 5 5600X, 64Гб) для изображений с наибольшим средним размером 2479×3589 пикселей составило 12,31±1,53 мс.

Ключевые слова: автоматическая обработка изображений документов, определение типа документа, динамическое программирование.

DOI: 10.14357/20790279250205 **EDN:** PXWRAG

Введение

Документы играют важную роль в современном мире. Без них нельзя устроиться на работу, путешествовать, водить машину, получить высшее образование и многое другое. При этом далеко не всегда используются физические носители, зачастую оказывается достаточно графического образа документа. Это стало сильно заметно при минимизации личных контактов во время пандемии COVID-19, особенно в делопроизводстве.

Для распознавания графических образов документов широко используются специализиро-

ванные системы автоматического распознавания изображений документов (СРИД) [1]. Они состоят из множества модулей, в рамках которых решаются различные задачи на стыке искусственного интеллекта, методов обработки изображений. Примеры прохождения изображения документа через модули СРИД проиллюстрирован рис. 1. Локализация изображения документа на входном изображении [2] предотвращает распознавание мусорных символов. После типизации документа [3] СРИД получит контекст, без которого нельзя верно интерпретировать информацию (например,



Рис. 1. Диаграмма типового процесса обработки документа в СРИД. Серым отмечен этап типизации, рассматриваемый в работе

какие поля на документе присутствуют, какие они имеют свойства). Знание типа документа дает много информации о правилах заполнения каждого поля. После локализации изменяемых реквизитов [4] есть возможность использовать свойства поля для более качественного распознавания. Распознавание реквизитов (текст [5], баркоды [6], чекбоксы [7]) выдает начальный результат обработки. После постобработки в результате распознавания не будет типичных ошибок системы. Например, исправляется результат распознавания нулей и букв «О» по контексту.

Контекст данной работы ограничен рассмотрением модуля типизации документа. Этот модуль является важным модулем СРИД, так как неверное определение типа документа на изображении, как видно из диаграммы на рисунке 1, ведет к нарушению работы всех последующих этапов работы системы. В работе рассматривается идентификация документа на изображении, полученного при помощи сканера, на заранее известном наборе типов документов, что является частным случаем типизации документа на изображении. Такой частный случай имеет прикладной характер. Например, при устройстве на работу в России требуется обязательный перечень документов [8], в который входят удостоверение личности (например, паспорт), трудовая книжка, документ об образовании, документы воинского учета, справка об отсутствии судимости и административных наказаний, подтверждение регистрации в системе индивидуального учета. Изучение сканированных документов также является важной прикладной задачей, например, для государственных учреждений и кадровых агентств неприемлема фотография документа, принимаются только сканы.

При описанных выше ограничениях важно быстро определять тип документа, так как это напрямую влияет на эффективность бизнес-процессов. Например, при массовом найме крупных компаний или в работе кадровых агентств, ежедневно обрабатывающих множество заявок, даже минимальная задержка в анализе каждого документа

приводит к значительным временным и финансовым потерям. Таким образом, при разработке методов стоит уделять особое внимание их производительности.

В работе предлагается метод идентификации типа документов на изображениях, полученных при помощи сканера, из заранее известного набора типов. В этом методе проекции производных входного изображения сопоставляются с набором референтных проекций при помощи алгоритма динамического выравнивания временных рядов. Описана интеграция предлагаемого метода с реальными методами компенсации угла изображения.

В разделе 1 рассматриваются работы, посвященные определению типа документа на изображении, и используемые алгоритмы; в разделе 2 ставится математическая задача идентификации типа документа на изображении для ограниченного набора типов; в разделе 3 описывается предлагаемый метод; в разделе 4 описывается план экспериментов; в разделе 5 представлены их результаты; затем, в разделе 6 анализируются результаты экспериментов, а в разделе «Заключение» подводятся итоги работы.

1. Литературный обзор

В настоящее время известно множество методов идентификации типа документа на изображении. Согласно работе [3], их можно разделить на три группы: а) на основе первично распознанного на изображении текста; б) на основе визуального содержания, например, логотипов или локальных особенностей изображения; в) на основе анализа структуры образа документа. К первой группе принадлежат такие методы, как анализ семантического векторного представления n -грамм на изображении [9], посимвольный анализ текста при помощи нейронных сетей [10], поиск ключевых слов документа [11] и многие другие. Методы, полагающиеся на анализ текста, сильно зависят от качества его первичного распознавания. Такое распознавание зачастую оказывается низкого качества, поскольку

нет возможности опереться ни на семантику, ни на геометрию документа. То есть ключевое слово может не быть обнаружено из-за ошибки распознавания, то же касается n -грамм и посимвольного анализа текста. При использовании методов из данной группы требуется повторное распознавание текста для улучшения качества распознавания, что приводит к лишним вычислениям.

Ко второй группе относится определение типа документа при помощи анализа дескрипторов локальных особенностей изображения [12]. В последнее время такой подход используется наиболее часто. Например, популярная открытая библиотека компьютерного зрения OpenCV [13] предоставляет множество алгоритмов для детекции и вычисления локальных особенностей. Использование в СРИД методов, анализирующих визуальное содержание, может требовать от пользователя проведения ряда предварительных работ, например, по созданию шаблонов типов документа, фильтрации зон с локальными особенностями изображений, их ретушированию и т.д. Так, в статье [3] предлагается маскирование на документе-примере зоны с изменяемыми реквизитами, на которых не будет вестись поиск локальных особенностей изображения. Такие зоны заменяются белыми прямоугольниками, поэтому в установлении корреспонденции будут участвовать дескрипторы точек, найденных только на неизменяемых частях документа. Заметим также, что выбор метода выделения локальных особенностей изображения является нетривиальной задачей [14].

К третьей группе относится анализ макета изображения документа с использованием скрытой марковской модели [15], фрагментация изображения на элементы и определение их типов (линии разграфки, статический текст, поля ввода) [16].

Также существуют методы, которые нельзя отнести к одной группе. Например, авторами работы [3] упоминается, что в статье [17] предложен метод, комбинирующий визуальные и текстовые дескрипторы.

Заметим, что при использовании нейронных сетей для решения задачи определения типа документа на изображении для обучения нужен репрезентативный и, возможно, большой объем данных, которого может не быть. При добавлении нового типа документа классификационная нейронная сеть требует переобучения. И, наконец, на мало-мощных устройствах нейросетевой подход может быть вычислительно затратным [14].

К методам, основанным на анализе структуры документа, также относится метод из доклада [18]. В докладе сформулирована идея использова-

ния в качестве описателя типа документа на изображении пары его проекций и сопоставления их с эталонными проекциями при помощи алгоритмов динамического программирования. Данная работа более детально рассматривает метод доклада [18] и развивает его идеи. Подобный метод затем также использовался в статье [19], посвященной сегментации регистрационных номеров автомобилей.

2. Постановка задачи и связанные определения

Приведем формальную постановку задачи определения типа документа. Пусть задано конечное множество рассматриваемых типов документов $T = \{t_1, t_2, \dots, t_N\}$. Пусть для входного изображения I зафиксирован способ g для построения его описателя и для каждого элемента из множества T известны эталонные описатели типов документов $E = \{e_1, e_2, \dots, e_N\}$. Пусть определен способ оценки близости $D(e(t_i), g(I))$, где $g(I)$ описатель для входного изображения, а $e(t_i)$ – эталонный описатель. Тогда, задача определения типа документа сводится к выбору наиболее близкого эталонного описателя к входному:

$$i_{best} = \underset{i \in \{1, \dots, N\}}{\operatorname{argmin}} D(e(t_i), g(I)), \quad (1)$$

В работе рассматриваются деловые документы, которые, как правило, являются представителями жестких структурированных документов. С формальным определением жесткого структурированного документа можно ознакомиться в работе Постникова В.В. [11]. Неформально, такие документы отличаются тем, что их статические элементы совпадают при наложении «на просвет» для всех образцов. Отличительной чертой класса деловых документов является их исходное проектирование для обработки с помощью ЭВМ, в результате чего они имеют множество таких графических элементов, как линии разграфки, решетки знакомест, таблицы, блоки текста, чекбоксы, штрих-коды и т.д. Показательные фрагменты типовых примеров документов такого класса приведены на рис. 2.

В предлагаемом методе сравниваются проекции изображений с использованием алгоритма динамической трансформации временной оси (далее ДТВО) [20]. Алгоритм позволяет найти соответствие между временными последовательностями, учитывая их возможные сдвиги и растяжения во времени. Сложность алгоритма составляет $O(m \times n)$, где m и n – размеры сравниваемых последовательностей.

При сканировании входные изображения зачастую оказываются наклонены [21]. Поэтому в реальной системе для использования предлагае-

Form 1040 Department of the Treasury—Internal Revenue Service U.S. Individual Income Tax Return 2022 OMB No. 1545-0074 IRS Use Only—Do not write or staple in this space.

Filing Status ☐ Single ☒ Married filing jointly ☐ Married filing separately (MFS) ☐ Head of household (HOH) ☐ Qualifying surviving spouse (QSS)

Check only one box. If you checked the MFS box, enter the name of your spouse. If you checked the HOH or QSS box, enter the child's name if the qualifying person is a child but not your dependent:

Your first name and middle initial Jason	Last name Clark	Your social security number 5 6 8 9 9 2 0 1 5
If joint return, spouse's first name and middle initial Olivia	Last name Clark	Spouse's social security number 4 7 8 0 1 4 7 8 3
Home address (number and street). If you have a P.O. box, see instructions. PO BOX 1002		Apt. no. 12
City, town, or post office. If you have a foreign address, also complete spaces below. LITCHFIELD		State MN
Foreign country name		ZIP code 55355-1003
Foreign province/state/county		Foreign postal code

Presidential Election Campaign Check here if you, or your spouse if filing jointly, want \$3 to go to this fund. Checking a box below will not change your tax or refund.
☐ You ☒ Spouse

(a)

cerfa

N° 12541 *02

feuille de soins - médecin

Art. R. 161-40 et suivants du Code de la sécurité sociale
Art. L. 212-1 du Code des pensions militaires d'invalidité et des victimes de la guerre

numéro de facture (facultatif)

date

PERSONNE RECEVANT LES SOINS ET ASSURÉ(E)

PERSONNE RECEVANT LES SOINS (la ligne "nom et prénom" est obligatoirement remplie par le médecin)

nom et prénom

nom de famille (de naissance) suivi du nom d'usage (facultatif et s'il y a lieu)

numéro d'immatriculation

date de naissance

ASSURÉ(E) (à remplir si la personne recevant les soins n'est pas l'assuré(e))

nom et prénom

nom de famille (de naissance) suivi du nom d'usage (facultatif et s'il y a lieu)

numéro d'immatriculation

ADRESSE DE L'ASSURÉ(E)

code de l'organisme de rattachement en cas de dispense d'avance des frais (à remplir par le médecin)

(b)

0000000000

первичный ☐ дубликат ☐

ЛИСТОК НЕТРУДОСПОСОБНОСТИ

продолжение листа нетрудоспособности №

Ф

И

О

(фамилия, имя, отчество нетрудоспособного)

(место работы-наименования организации)

Основное ☐ По совместительству ☐ №

001 234 567 891

(фамилия, инициалы врача)

№ истории болезни

Дата выдачи

расписка получателя

(c)

Рис. 2. Примеры фрагментов типовых деловых документов с жесткой структурой: а) налоговая форма США; б) французская форма лечения; с) российский листок нетрудоспособности

мого метода потребуются предварительно автоматически нормализовать входные изображения. В данной работе в качестве метода нормализации используется метод, опубликованный в работе [22]. Этот метод базируется на анализе образа быстрого преобразования Хафа (далее – БПХ) [23] и отличается повышенной производительностью, что явля-

ется важным фактором в контексте данной работы. При этом, вычисление БПХ в нем осуществляется с помощью алгоритма Брейди-Ёна [24].

Введем определения, которые будут использоваться далее. Для сравнения проекций входного изображения с референтными, последние должны быть уже подготовлены. Назовем **эталонным изо-**

теля g возьмем пару проекций производных изображения – проекцию на ось Oy производной изображения по x и проекцию на ось Ox производной изображения по y . В докладе [18] при решении этой задачи упоминались два пункта, на которые стоит обратить внимание. Во-первых, если изображение не нормализовано, то есть линии таблиц не параллельны осям координат, то яркостные проекции перестают отражать табличные свойства документа и становятся непригодными для типизации. Пример различия проекций приведен на рис. 3. Во-вторых, одного изображения для типизации может быть недостаточно, так как на документе присутствуют не только статические элементы, но и изменяемые поля, которые вносят свой вклад в проекции. Две проекции двух различных нормализованных изображений документа одного типа не будут равны. То есть, нужно некоторое количество изображений для построения эталонных проекций и их векторов отклонений. Первая проблема решается использованием быстрого преобразования Хафа и выбором нужной строки – проекции. Для решения второй мы воспользуемся вектором отклонений.

Теперь построим эталонные проекции типа документа. Если изображение нормализовано, то проекции считаем, суммируя пиксели на изображении проекции вдоль осей Ox и Oy , однако в реальных условиях изображения редко бывают нормализованными, поэтому воспользуемся быстрым преобразованием Хафа. Алгоритм вычисления проекции для ненормализованного изображения будет описан для проекции по оси Ox , для оси Oy аналогично.

Предварительно обрабатываем каждое эталонное изображение рассматриваемого типа документа. Сначала масштабируем изображение ε так, чтобы его ширина была равна w , а высота изменилась пропорционально, получаем изображение ε' . Применяем к ε' фильтр Гаусса Δ [25] с параметром σ_{gauss} для снижения шума на изображении, получаем изображение $\Delta(\varepsilon')$. Затем вычисляем производную изображения $\Delta(\varepsilon')$ по y [25], получаем изображение Dx . Вычисляем по производной Dx ее БПХ-образ по горизонтальным направлениям [26], получаем изображение $FHT(Dx)$. Вычисляем индекс i строки БПХ-образа, который отвечает за проекцию вдоль горизонтальных линий таблиц [22]. Наконец, по индексу i получаем строку БПХ-образа $FHT(Dx)[i]$ – необходимую проекцию на ось Ox .

Теперь для каждого эталонного изображения построена соответствующая ему проекция. Для построения эталонной проекции используем алгоритм, представленный на рис. 4.

Algorithm 1 Алгоритм построения эталонной проекции

Require: Массив π проекций эталонных изображений из n элементов
Ensure: Эталонная проекция

```

1:  $changed \leftarrow \{\}$   $\triangleright$  Множество измененных индексов массива
2:  $a \leftarrow 0$   $\triangleright$  Индекс изменяемого вектора
3: while  $changed.size() + 1 \neq n$  do
4:   for  $i \leftarrow 0$  to  $n$  do
5:     if  $i \in changed$  then
6:       continue
7:     end if
8:      $min\_dist \leftarrow +\infty$ 
9:     for  $j \leftarrow i + 1$  to  $n$  do
10:      if  $j \in changed$  then
11:        continue
12:      end if
13:       $d \leftarrow DTW(\pi[i], \pi[j])$ 
14:      if  $d < min\_dist$  then
15:         $min\_dist \leftarrow d$ 
16:         $changed.insert(j)$ 
17:         $a \leftarrow j$ 
18:      end if
19:    end for
20:     $\pi[a] \leftarrow (\pi[a] + \pi[i]) / 2$   $\triangleright$  Поэлементная операция с учетом ДТВО
21:  end for
22: end while
23: return  $\pi[a]$ 

```

Рис. 4. Алгоритм построения эталонной проекции

После алгоритма на рис. 4 получена эталонная проекция, однако не учитывается, что на документах есть изменяемые поля. Для решения этой проблемы построим вектор, размерность которого равна размерности уже построенного вектора, отвечающий за отклонения, алгоритм приведен на рис. 5.

Таким образом, строится эталонная проекция типа документа и вектор отклонений для каждого типа. Заметим, что вектор отклонений указывает на возможное положение изменяемых реквизитов.

Теперь определим к какому типу принадлежит документ на изображении:

1. Предварительно обрабатываем изображение, как описано выше, и получаем его проекцию.
2. Для каждой пары эталонных проекций типа документа вычисляем расстояние до эталонных проекций, добавляя весовую функцию в алгоритм ДТВО, причем $Dist(0, 0) = \frac{|e_0 - s_0|}{\sqrt{\sigma_0^2 + \lambda^2}}$.

$$Dist(i, j) = \frac{|e_i - s_j|}{\sqrt{\sigma_i^2 + \lambda^2}} + \min \begin{cases} P_{del} + Dist(i - 1, j) \\ Dist(i - 1, j - 1) \\ P_{ins} + Dist(i, j - 1), \end{cases} \quad (2)$$

где σ_i – i -ая координата соответствующего вектора отклонений, λ – настраиваемый параметр, e_i – i -ая координата эталонной проекции, s_j – j -ая координата проекции входного изображения, P_{del} и P_{ins} – величина штрафов за вставку в эталонную и входную проекцию соответственно.

3. Находим минимум по значениям функции критерия. Тип, значение до эталонных проекций которого наименьшее, является типом документа на входном изображении по формуле 1.

Построенный метод определения типа документа на произвольно повернутых изображениях

Algorithm 2 Алгоритм построения вектора отклонений

Require: Массив π проекций эталонных изображений из n элементов, эталонная проекция S , размера m
Ensure: Вектор отклонений

```

1:  $V \leftarrow \mathbf{0}$   $\triangleright V$  – нулевой вектор размера  $m$ 
2: for  $i \leftarrow 0$  to  $n-1$  do
3:   for  $j \leftarrow 0$  to  $m-1$  do
4:      $k \leftarrow I_j$   $\triangleright$  Координата вектора  $\pi[i]$ , соответствующая координате  $j$  эталонной проекции  $S$ 
5:      $\sigma \leftarrow (|\pi[i][k] - S[j]|)/2$ 
6:     if  $\sigma > V[j]$  then
7:        $V[j] \leftarrow \sigma$ 
8:     end if
9:   end for
10: end for
11: return  $V$ 

```

Рис. 5. Алгоритм построения вектора отклонений

является полностью автоматическим. Для автоматизированных систем при построении эталонной проекции угол наклона эталонных изображений может быть компенсирован вручную.

4. План эксперимента

Для экспериментов был создан корпус данных K из сканов восьми жестких типов документов. В корпус K вошли такие типы документов как T_1 – второй регистрационный бланк; T_2 – третий регистрационный бланк; T_3 – четвертый регистрационный бланк; T_4 – сведения о показаниях квартирных приборов (рис. 3); T_5 – листок нетрудоспособности; T_6 – страховой полис; T_7 – заявление о выборе инвестиционного портфеля; T_8 – заявление на предоставление основной банковской карты.

Количество изображений каждого типа документа приведено в табл. 1. Для каждого типа документа было случайным образом отобрано 15 изображений для построения эталонных проекций.

Для каждого изображения из K была вручную проведена процедура геометрической нормализации, а именно компенсация угла наклона изображения документа, в результате которой опорные линии на его изображении оказались параллельны осям координат. Такое изображение далее будем называть **идеально нормализованным**. Идеально нормализованное изображение нужно для сравнения работы алгоритма в идеальных условиях и в условиях автоматического поворота изображения.

Для оценки точности определения типа документа Q используем функцию:

$$Q = \frac{\sum_i |c_t^i - c_m^i|}{\text{Card}(c_m)}, \quad (3)$$

где c_t^i – тип документа на изображении, который определил метод, c_m^i – истинное значение типа, $\text{Card}(c_m)$ – мощность множества всех обработанных изображений, $[\cdot]$ – нотация Айверсона.

В экспериментах $\lambda=1$ в формуле 2, в качестве оценки близости между проекциями входного изображения $P = \{p_x, p_y\} = R(I)$ и эталонными проекциями $e(t_i) = \{p_x^0, p_y^0\}$ используется величина (4):

$$D(P, e(t_i)) = \sqrt{\frac{DTW(p_x, p_x^0) + DTW(p_y, p_y^0)}{|p_x| + |p_y| + |p_x^0| + |p_y^0|}}, \quad (4)$$

где величина $DTW(a, b)$ – обозначает расстояние ДТВО, то есть правый нижний заполненный элемент матрицы $\text{Dist}(i, j)$ из формулы 2 для векторов a и b .

Параметры w (ширина изображения) и σ_{gauss} равны 128 и 1.5 соответственно.

Для эксперимента на идеально нормализованных изображениях производные и БПХ-образы не строятся. Так как заранее известно, что проекции на оси координат совпадают с нужными нам проекциями, посчитаем их, сразу просуммировав пиксели вдоль осей координат.

5. Вычислительный эксперимент

Обозначим предлагаемый метод М1, а референтный метод из статьи [12] – М2. Результаты работы метода М1 на идеально нормализованных изображениях приведены в табл. 2а. Согласно формуле 3, точность типизации $Q=99.79\%$. Результаты работы метода М1 для изображений, угол поворота, которых априори не известен, приведены в табл. 2б. На них точность типизации $Q = 99.76\%$.

Результаты определения типа документа на изображении для идеально нормализованных изображений методом М2 приведены в таблице 3а, точность типизации $Q = 99.41\%$. 12 изображениям, содержащим документ типа T_5 , тип документа не был присвоен, реализация метода выдала ответ «untypified». Результаты определения типа документа на произвольных изображениях методом М2 приведены в табл. 3б, результат полностью совпадает с результатами, приведенными в 3а.

Время работы для каждого типа документа приведено в табл. 4. Под t_{type} понимается среднее время определения типа документа на идеально нормализованном изображении методом М1, то есть время вычисления ДТВО расстояний до эталонных

Табл. 1

Тип документа и соответствующее ему количество изображений.

Тип документа	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
Количество изображений	462	472	958	329	246	193	279	61

Табл. 2

Типизация документов методом М1: а) матрица ошибок эксперимента на идеально нормализованных изображениях; б) матрица ошибок эксперимента на повернутых изображениях. Пустым ячейкам соответствуют нули

	а								б							
	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈
T ₁	447								447							
T ₂		467								467						
T ₃			942								942					
T ₄				314								314				
T ₅					230								230			
T ₆						172		6					6	172		
T ₇							264							1	263	
T ₈								46								46

Табл. 3

Типизация документов методом М2: а) матрица ошибок на идеально нормализованных изображениях; б) матрица ошибок на повернутых изображениях. Пустым ячейкам соответствуют нули, U обозначает «untypified».

	а									б								
	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	U	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	U
T ₁	447									447								
T ₂		467									467							
T ₃			942									942						
T ₄				314									314					
T ₅				4	213	1			12				4	213	1			12
T ₆						178									178			
T ₇							264									264		
T ₈								46									46	

проекций и выбор минимума из этих расстояний. Под t_{total} понимается среднее общее время метода М1, то есть время изменения размера не нормализованного изображения, время вычисления производных и БПХ-образов, время вычисления проекций и время определения типа. Под t_{kps} понимается среднее время определение типа документа на идеально нормализованном изображении метода М2.

Все эксперименты проводились на ПК с процессором AMD Ryzen 5 5600X (6 ядер, 12 потоков) и 64Гб оперативной памяти. Программа написана на языке C++ с компилятором GCC 12.4.1.

6. Обсуждение

Согласно табл. 2а, все ошибки типизации метода М1 возникают на изображениях документа типа T_6 , тип которых неверно определяется как T_8 . Анализ ошибок показал, что при этом тип T_6 является второй альтернативой. Ошибка функции-критерия на этих изображениях лежит в диапазоне от 2.52% до 8.46%. Заметим, что в пяти из шести

случаях расстояние ДТВО до каждой из эталонных проекций T_6 больше, чем до эталонных проекций T_8 , только в одном случае расстояние до проекции по Оу у T_6 меньше.

Согласно табл. 2б, тип документа T_6 на шести изображениях неверно определен как T_5 и один типа T_7 определен как T_6 . Анализ ошибок показал, что только одно изображения, на котором неверно определен тип, совпадает с ошибочным изображением для идеально нормализованного случая. В данном случае для трех изображений тип T_6 не является второй альтернативой для тип документа, им оказывается тип T_2 . Аналогично для изображения с документом типа T_7 ; второй альтернативой также является T_2 . У всех семи изображений был неверно определен угол поворота, значит, неверно найдена проекция.

Отметим, что изображение, на котором возникает ошибка в случае идеальной нормализации и с поворотом, имеет множественные «паразитные» горизонтальные зачеркивания, которые меняют его структуру. Часть изображения приведена на рис. 6.

Табл. 4

Время работы метода для каждого типа документа.

Тип документа	Средний размер изображения, пэл	t_{type} , мс	t_{total} , мс	t_{kps} , мс
T_1	1664x2338	3.51 ± 0.13	9.37 ± 0.29	86.14 ± 15.35
T_2	1664x2338	3.73 ± 1.04	9.99 ± 0.54	107.44 ± 13.31
T_3	2338x1664	2.08 ± 0.85	6.77 ± 0.58	205.86 ± 28.82
T_4	1653x932	1.74 ± 0.23	5.70 ± 0.77	40.30 ± 9.29
T_5	2474x2954	3.16 ± 0.13	10.86 ± 1.90	190.06 ± 52.22
T_6	2479x3589	3.49 ± 0.20	12.31 ± 1.53	152.16 ± 16.60
T_7	2484x3503	3.46 ± 0.18	10.07 ± 0.99	81.69 ± 10.09
T_8	1655x2331	3.69 ± 0.08	9.77 ± 0.19	151.04 ± 19.73

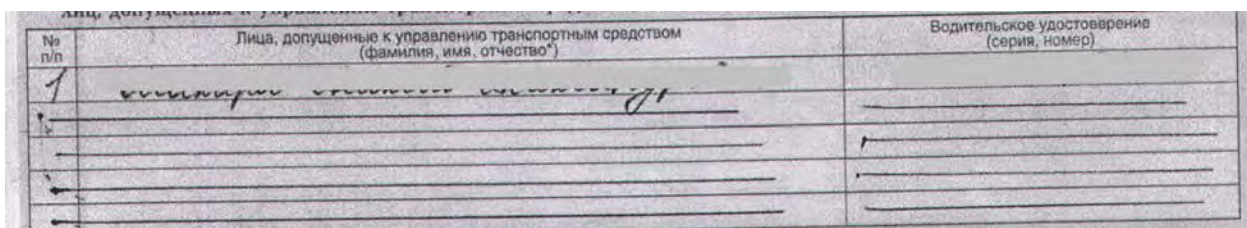


Рис. 6. Часть изображения типа T_6 из корпуса данных K, на котором множественные зачеркивания изменяют его структуру. Серым цветом замаскирована конфиденциальная информация

Сравнение результатов метода M1 и метода M2, приведенных на таблицах 2 и 3, демонстрирует, что метод M1 лучше в обоих случаях для типа T_5 , однако тип T_6 в методе M2 определяется безошибочно. Средняя точность типизации метода M1 для нормализованных изображений выше на 0.38%, для произвольных – на 0.35%, а количество ошибок метода M2 сокращается на 64.71% и 58.82% соответственно.

Согласно табл. 4, метод M2 для любого типа документа работает медленнее метода M1 в среднем в 10 раз. Общее время работы t_{total} в среднем больше времени работы определения типа t_{type} в 3 раза.

Заключение

В работе предложен метод определения типа делового документа на сканированном изображении из ограниченного набора типов. Метод основан на сопоставлении проекций производных изображений документа при помощи алгоритма динамической трансформации временной оси.

Предложенный метод не нуждается в бинаризации и предварительной обработке шаблона документа пользователем (например, выделение ключевых слов, геометрических примитивов и т.д.). Однако требуется компенсация угла наклона изображения. В работе была показана интеграция предложенного метода с существующим методом нормализации изображения на базе БПХ.

На корпусе данных, состоящем из восьми типов документов и 3008 изображений, предложенный метод показал высокую точность типизации – 99.79% на вручную нормализованных изображениях и 99.76% на произвольных изображениях. В сравнении метод, основанный на анализе дескрипторов локальных особенностей изображения, показал точность определения типа 99.41%. Более того, предложенный метод работает в среднем в 10 раз быстрее чем сравниваемый метод на нормализованных изображениях, что является важным аспектом для данной задачи. А именно, для типа документа с наибольшим средним размером изображений 2479x3589 пикселей среднее время работы метода составляет $12,31 \pm 1,53$ мс.

Литература

1. Arlazarov V.V., Andreeva E.I., Bulatov K.B., Nikolaev D.P., Petrova O.O., Savelev B.I., Slavin O.A. Document image analysis and recognition: A survey. // Computer Optics. 2022. 46(4) p. 567–589. DOI: 10.18287/2412-6179-CO-1020.
2. Skoryukina N.S., Arlazarov V.V., Nikolaev D.P., Faradjev I.A. Efficient Location and Identification of Documents in Images. Патент № US11574492B2, 02.09.2020 // United States Patent. С. 1-25.
3. Awal A.M., Ghanmi N., Sicre R., Furon T. Complex document classification and localization application on identity document images. // In 2017 14th IAPR

- International Conference on Document Analysis and Recognition (ICDAR). 2017. 426–431. IEEE. DOI:10.1109/ICDAR.2017.77.
4. *Gayer A.V., Arlazarov V.V.* Muldt: Multilingual ultra-lightweight document text detection for embedded devices. // IEEE Access. 2024. 12. p. 170530–170540. DOI: 10.1109/ACCESS.2024.3474616.
 5. *Bahi H.E., Zatni A.* Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network. // Multimedia Tools and Applications. 2019. 78(18) p. 26453–26481. DOI:10.1007/s11042-019-07855-z.
 6. *Славин О.А., Федоров Г.О.* Об использовании штрих-кодирования и специализированных устройств в корпоративном электронном документообороте // Труды ИСА РАН. 2003. Т. 4. С. 185–1972.
 7. *Shengnan Z., Shanlei Y., Lianqiang N.* Automatic recognition method for checkbox in data form image. // In 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation. 2014. page 159–162. IEEE. DOI:10.1109/ICMTMA.2014.42.
 8. Какие документы могут потребовать при устройстве на работу? [Электронный ресурс] // КонсультантПлюс. URL: https://www.consultant.ru/edu/student/consultation/dokumenty_ustroystvo_na_rabotu/ (дата обращения: 01.05.2025)
 9. *Seifollahi S., Piccardi M., Jolfaei A.* An embedding-based topic model for document classification. // ACM Trans. Asian Low-Resour. Lang. Inf. Process. 20(3). 1–13. DOI:10.1145/3431728.
 10. *Xiao Y., Cho K.* Efficient character-level document classification by combining convolution and recurrent layers. // arXiv 2016. DOI:10.48550/arXiv.1602.00367.
 11. *Постников В.В.* Автоматическая идентификация и распознавание структурированных документов: автореферат диссертации на соискание ученой степени кандидата технических наук. // ИСА РАН 2001.
 12. *Skoryukina N., Arlazarov V., Nikolaev D.* Fast method of ID documents location and type identification for mobile and server application. // In 2019 International Conference on Document Analysis and Recognition (ICDAR) 2019. pages 850–857. IEEE. DOI:10.1109/ICDAR.2019.00141.
 13. Feature detection and description. [Электронный ресурс] // Open Source Computer Vision. – URL: https://docs.opencv.org/4.x/db/d27/tutorial_py_table_of_contents_feature2d.html. (дата обращения: 01.05.2025)
 14. *Skoryukina N.S., Tropin D.V., Shemiakina Y.A., Arlazarov V.V.* Document localization and classification as stages of a document recognition system. // Pattern Recognit. Image Anal. 2023. 33(4). 699–716. DOI: 10.1134/S1054661823040430.
 15. *Hu J., Kashi R., Wilfong G.* Document classification using layout analysis. // In Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99. pages 556–560. IEEE, 1999. DOI:10.1109/DEXA.1999.795245.
 16. *Постников В.В.* Формальный подход к задаче идентификации графических образов структурированных документов. // ИТиВС. 1999.— (4):280–299.
 17. *Rusin~ol M., Frinken V., Karatzas D., Bagdanov A. D., Llado's J.* Multimodal page classification in administrative document image streams. // International Journal on Document Analysis and Recognition (IJ DAR). 2014. 17(4). 331–341. DOI:10.1007/s10032-014-0225-8.
 18. *Безматерных П., Николаев Д., Постников В.* Метод идентификации типа документа по структуре проекций его изображения на координатные оси. // ИТиС. 2008. с. 498–501. ISBN 978-59-01158-08-0.
 19. *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. 32(1). С. 50–59. DOI: 10.7868/S0235009218010080.
 20. *Sakoe H., Chiba S.* Dynamic programming algorithm optimization for spoken word recognition. // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1(26):43–49. DOI:10.1109/TASSP.1978.1163055.
 21. *Hull J.J.* Document images skew detection: survey and annotated bibliography. // World Scientific. 1998. p. 40–64. DOI:10.1142/9789812797704_0003.
 22. *Безматерных П.В.* Нормализация изображения текста с помощью быстрого преобразования Хафа. // ИТиВС. 2024. (4):3–16. DOI: 10.14357/20718632240401.
 23. *Bezmaternykh P.V., Nikolaev D.P., Arlazarov V.L.* High-performance digital image processing. // Pattern Recognit. Image Anal. 2023. 33(4). 743–755. DOI: 10.1134/S1054661823040090.
 24. *Brady M.L., Yong W.* Fast parallel discrete approximation algorithms for the radon transform. // In Proceedings of the fourth annual ACM symposium on Parallel algorithms and architectures. pages 91–99. ACM. DOI:10.1145/140901.140911.
 25. *Gonzalez R.C., Woods R. E., Masters B.R.* Digital image processing, third edition. 14. 029901.

26. Алиев М.А., Кунина И.А., Николаев Д.П., Полевой Д.В. О практических аспектах вычисления Хаф-образа алгоритмом Брейди-Ёна. // Информационные процессы. 2023. 23(2). с. 250–273. DOI: 10.53921/18195822_2023_23_2_250.

Максимова Таисия Романовна. ООО «Смарт Энджинс Сервис», г. Москва, Россия. Программистка. Область научных интересов: распознавание текста, классификация типа документа на изображении, системы распознавания документов. E-mail: t.maksimova@smartengines.com (ответственная за переписку).

Безматерных Павел Владимирович. Федеральный исследовательский центр «Информатика и управление» Российской академии наук, г. Москва, Россия. Программист первой категории. ООО «Смарт Энджинс Сервис», г. Москва, Россия. Научный сотрудник-программист. Область научных интересов: обработка изображений документов, распознавание штрихкодов. E-mail: bezmaternyh@isa.ru

Classification of a scanned document type using the dynamic time warping method

T.R. Maximova^I, P.V. Bezmaternykh^{I,II}

^I Smart Engines Service LLC, Moscow, Russia

^{II} Federal Research Center “Computer Science and Control”, Russian Academy of Sciences, Moscow, Russia

Abstract. The work addresses an important problem in the field of automatic document image recognition: determining the type of a scanned document from a predefined set of possible types. The proposed document classification method compares parallel projections of the input image with reference projections of templates from the target set, which can be generated using just a few document image samples. The matching is performed using a dynamic time warping algorithm. The classification method requires neither prior binarization of the sample, nor keyword extraction or recognition, nor detection of geometric primitives. However, it does require preliminary image deskewing. Experiments were conducted on a manually normalized dataset of business documents comprising eight distinct types, achieving a classification accuracy of 99.79%. For the same images normalized automatically, the accuracy reached 99.76%. For the document type with the largest average image size (2479x3589 pxs), the average processing time is 12.31 ± 1.53 ms on a PC with an AMD Ryzen 5 5600X CPU, 64GB RAM.

Keywords: automatic document image processing, document type classification, dynamic programming.

DOI: 10.14357/20790279250205 **EDN:** PXWRAG

References

1. Arlazarov V.V., Andreeva E.I., Bulatov K.B., Nikolaev D.P., Petrova O.O., Savelev B.I., Slavin O.A. Document image analysis and recognition: A survey. *Computer Optics*. 2022; 46(4):567–589. DOI: 10.18287/2412-6179-CO-1020.
2. Skoryukina N.S., Arlazarov V.V., Nikolaev D.P. and Faradjev I.A. “Efficient Location and Identification of Documents in Images,” United States Patent. Patent № US11574492B2, 02.09.2020. P. 1-25, 2023.
3. Awal A.M., Ghanmi N., Sicre R., Furon T. Complex document classification and localization application on identity document images. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). 2017; 426–431. IEEE. DOI:10.1109/ICDAR.2017.77.
4. Gayer A.V., Arlazarov V.V. Muldt: Multilingual ultra-lightweight document text detection for embedded devices. *IEEE Access*. 2024; 12:170530–170540. DOI: 10.1109/ACCESS.2024.3474616.
5. Bahi H.E., Zatni A. Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network. *Multi-media Tools and Applications*. 2019; 78(18):26453–26481. DOI:10.1007/s11042-019-07855-z.
6. Slavin O.A. and Fedorov G.O. “Ob ispolzovanii shtrikh-kodirovaniya i spetsializirovannykh ustroystv v korporativnom elektronnom dokumentooborote,” *Trudy ISA RAN (Proceedings of ISA RAS)*, vol. 4, pp. 185-197, 2003.
7. Shengnan Z., Shanlei Y., Lianqiang N. Automatic recognition method for checkbox in data form

- image. In 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation. 2014; page 159–162. IEEE. DOI:10.1109/ICMTMA.2014.42.
8. ConsultantPlus. What documents may be required when applying for a job? 2024. (In Russ.) URL: https://www.consultant.ru/edu/student/consultation/dokumenty_ustroystvo_na_rabotu/.
 9. Seifollahi S., Piccardi M., Jolfaei A. An embedding-based topic model for document classification. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 20(3):1–13. DOI:10.1145/3431728.
 10. Xiao Y., Cho K. Efficient character-level document classification by combining convolution and recurrent layers. arXiv. 2016. DOI:10.48550/arXiv.1602.00367.
 11. Postnikov V.V. “Avtomaticheskaya identifikatsiya i raspoznavanie strukturirovannykh dokumentov”. Dissertatsiya, pp. 1-126, 2001.
 12. Skoryukina N., Arlazarov V., Nikolaev D. Fast method of ID documents location and type identification for mobile and server application. In 2019 International Conference on Document Analysis and Recognition (ICDAR). 2019; pages 850–857. IEEE. DOI:10.1109/ICDAR.2019.00141.
 13. Open Source Computer Vision. Feature detection and description. URL: https://docs.opencv.org/4.x/db/d27/tutorial_py_table_of_contents_feature2d.html.
 14. Skoryukina N.S., Tropin D.V., Shemiakina Y.A., Arlazarov V.V. Document localization and classification as stages of a document recognition system. Pattern Recognit. Image Anal. 2023; 33(4):699–716. DOI: 10.1134/S1054661823040430.
 15. Hu J., Kashi R., Wilfong G. Document classification using layout analysis. In Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99, pages 556–560. IEEE, 1999. DOI:10.1109/DEXA.1999.795245.
 16. Postnikov V.V. Formalnyy podkhod k zadache identifikatsii graficheskikh obrazov strukturirovannykh dokumentov. ITiVS, (4):280–299, 1999.
 17. Rusin~ol M., Frinken V., Karatzas D., Bagdanov A. D., Llado's J. Multimodal page classification in administrative document image streams. International Journal on Document Analysis and Recognition (IJ DAR). 2014; 17(4):331–341. DOI:10.1007/s10032-014-0225-8.
 18. Bezmaternykh P., Nikolaev D., Postnikov V. Metod identifikatsii tipa dokumenta po strukture proektsiy ego izobrazheniya na koordinatnye osi. ITaS. 2008; c. 498–501, 2008. ISBN 978-59-01158-08-0.
 19. Povolotskiy M.A., Kuznetsova E.G., Utkin N.V., Nikolaev D.P. Segmentation of vehicle registration plates based on dynamic time warping. Sensory systems. 2018; 32(1):50–59. DOI: 10.7868/S0235009218010080.
 20. Sakoe H., Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1(26):43–49. DOI:10.1109/TASSP.1978.1163055.
 21. Hull J.J. Document image skew detection: survey and annotated bibliography. World Scientific. 1998. P. 40–64. DOI:10.1142/9789812797704_0003.
 22. Bezmaternykh P.V. Text Image Normalization Using Fast Hough Transform. ITiVS. 2024; (4):3–16. DOI: 10.14357/20718632240401.
 23. Bezmaternykh P.V., Nikolaev D.P., Arlazarov V.L. High-performance digital image processing. Pattern Recognit. Image Anal. 2023; 33(4):743–755. DOI: 10.1134/S1054661823040090.
 24. Brady M.L., Yong W. Fast parallel discrete approximation algorithms for the radon transform. In Proceedings of the fourth annual ACM symposium on Parallel algorithms and architectures. pages 91–99. ACM. DOI:10.1145/140901.140911.
 25. Gonzalez R.C., Woods R.E., Masters B.R. Digital image processing, third edition. 14(2):029901.
 26. Aliev M.A., Kunina I.A., Nikolaev D.P., Polevoy D.V. On the practical aspects of computing the Hough image by the Brady-Yong algorithm. Informatsionnye protsessy. 2023; 23(2):250–273. DOI: 10.53921/18195822_2023_23_2_250.

Taisiya R. Maximova. Smart Engines LLC, Moscow, Russia. Topics of interest: document type classification in images, document recognition systems. E-mail: t.maksimova@smartengines.com

Pavel V. Bezmaternykh. Federal Research Center «Computer Science and Control» of Russian Academy of Sciences, Moscow, Russia. Smart Engines LLC, Moscow, Russia. Topics of interest: document image analysis. E-mail: bezmaternyh@isa.ru