

Нормализация изображения текста с помощью быстрого преобразования Хафа

П. В. Безматерных^{I,II}

^I Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, Россия

^{II} ООО «Смарт Энджинс Сервис», Москва, Россия

Аннотация. В работе приведено решение для двух классических задач геометрической нормализации цифрового образа текста: компенсация глобального угла наклона документа и устранение локальных наклонов его текстовых фрагментов. Для обеих задач, отличающихся видом геометрических искажений, решение построено на базе единого метода анализа образа быстрого преобразования Хафа. Проведена конкретизация данного метода и предложены два алгоритма для решения указанных задач, а также проведена их апробация: для задачи нормализации наклона текста – на множестве как известных корпусов данных, так и на специально собранном и опубликованном корпусе кириллических фрагментов *KRUS*; для задачи нормализации наклона документа – на популярном корпусе DISEC. Показано, что отличительной особенностью предложенного метода является высокое быстродействие с возможностью обработки большого диапазона углов, а сам метод может быть успешно применен в системах автоматической обработки изображений документов.

Ключевые слова: нормализация изображений, быстрое преобразование Хафа, анализ изображений документов.

DOI 10.14357/20718632240401 EDN VOYYXQ

Введение

В работе рассматриваются две классические задачи из области автоматической обработки изображений документов, связанные с геометрической нормализацией цифрового образа документа: D_{skew} – компенсация наклона всего документа и D_{slant} – исправление наклонного начертания его отдельных текстовых фрагментов.

Входное изображение документа, как правило, получается путем сканирования или фотографирования физического носителя. Данный процесс подвержен влиянию как человеческого, так и технического факторов, что часто приводит к геометрическим искажениям цифрового образа документа. Наклон документа является одним из наиболее распространенных примеров

таких искажений, а его компенсация – ключевым этапом автоматической обработки изображения документа [1]. После нормализации изображения все последующие стадии его анализа, такие как выделение текстовых фрагментов и их обработка модулями оптического распознавания символов, могут быть существенно упрощены. На Рис. 1 приведено изображение документа до компенсации его наклона (слева) и после (по центру).

Точность распознавания текстового фрагмента существенно зависит от точности процедуры его сегментации на отдельные символы. Если шрифтовое начертание символов фрагмента является стандартным, то для его сегментации можно использовать модель вертикальных разрезов. Данная одномерная модель ранее



Рис. 1. Цифровой образ паспорта РФ: слева – до компенсации наклона; по центру – после; справа – пример ID карты Болгарии с текстовыми фрагментами наклонного начертания

отлично зарекомендовала себя при решении задачи сегментации печатных символов [2; 3]. При этом она не потеряла своей актуальности и сейчас [4]. Чтобы ею можно было воспользоваться для фрагментов с нестандартными начертаниями, их необходимо предварительно нормализовать. К нестандартным начертаниям относятся: (а) механически наклоненные (англ. *oblique*), (б) курсивные (англ. *italic*), (в) рукописные фрагменты. На Рис. 2 приведены примеры реальных входных данных с нестандартным начертанием, а также ожидаемый результат их нормализации.

Успешное распознавание текстовых фрагментов с указанными начертаниями является существенным преимуществом при выборе системы распознавания для промышленного использования. Так, в задаче автоматического ввода паспортов гражданина РФ нередко встречаются документы, заполненные от руки, а идентификационные карты Болгарии изначально содержат атрибуты, напечатанные с наклонным начертанием (Рис. 1, справа). Наличие нестандартного начертания особенно сильно влияет на процедуру сегментации фрагмента на отдельные символы [3].

Для решения задачи D_{skew} к изображению ненормализованного документа применяется преобразование поворота R (англ. *rotate*), а в задаче D_{slant} к изображениям текстовых фрагментов применяется преобразование скоса S (англ. *slant*). Эти преобразования определяются матрицами:

$$R(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, S(\alpha) = \begin{bmatrix} 1 & -\tan(\alpha) \\ 0 & 1 \end{bmatrix} \quad (1)$$

Таким образом, обе задачи D_{skew} и D_{slant} сводятся к автоматическому определению угла наклона α . Для решения указанных задач зачастую используется анализ результата преобразования Хафа (далее – ПХ), которое можно рассматривать как дискретную версию интегрального преобразования Радона (далее – ДПР) [5]. Вычисление данного преобразования сопряжено с высокими вычислительными затратами (для квадратного изображения $I^{N \times N}$ требуется порядка $O(N^3)$ суммаций), поэтому вопросам эффективного построения ДПР посвящено много внимания.

В данной работе предлагается метод решения обеих задач, использующий быструю аппроксимацию вычисления ДПР с помощью алгоритма



Рис. 2. Координатный скос для нестандартных начертаний (верхний ряд) и результат нормализации (нижний ряд)

Брейди-Ёна [5]. Данная аппроксимация также известна как быстрое преобразование Хафа (далее – БПХ) [6]. Возможность применения БПХ для решения задачи D_{skew} была, по-видимому впервые, предложена в упомянутой работе [6]. К сожалению, в ней не приведен ни алгоритм, ни какие-либо характеристики метода, базирующегося на данной аппроксимации. И хотя оценка неточности быстрой аппроксимации ДПР исследована теоретически [7], вопрос насколько сильно данная неточность влияет на результаты в практических приложениях является открытым. Таким образом, исследование скоростных и точностных характеристик предлагаемого подхода является важным для установления границ применимости методов определения угла наклона в рамках промышленных систем автоматического анализа изображений документов. В тоже время, для задачи определения угла наклона текстового фрагмента решений с использованием БПХ выявить не удалось. Поэтому целью данной работы является синтез алгоритмов определения углов наклона для рассматриваемых задач и проведение вычислительных экспериментов на релевантных массивах данных для установления границ их практической применимости.

Далее приводится краткий обзор подходов к решению рассматриваемых задач. После чего, разбирается предлагаемый метод, особенности используемой аппроксимации, и, наконец, приводится два соответствующих алгоритма. Затем излагаются результаты проведенных вычислительных экспериментов на релевантных корпусах данных и приводится заключение работы.

Обзор литературы

Важность задач геометрической нормализации изображений текста была осознана с появлением первых систем оптического распознавания символов. По тематике задачи D_{skew} написано множество обзорных работ. Из них стоит выделить монументальное исследование

Дж. Хилла, опубликованное еще в 1996 году [8]. В нем приводится анализ более 50 работ по теме, описывающих свыше двух десятков соответствующих алгоритмов. Для каждого из них приведено краткое описание, а также выявлены ключевые идеи, которые были сгруппированы автором в четыре класса: (а) анализ параллельной проекции (далее – ПП), (б) анализ распределения особенностей, (в) анализ ориентированных компонент, (г) анализ результата преобразования Хафа.

Основная идея группы (а) заключается в анализе одной или нескольких ПП входного изображения I на координатные оси [9]. В таком случае, задача D_{skew} сводится к последовательному повороту исходного изображения на углы из списка и выбору лучшего проекционного профиля. Для возможности ранжирования оценок полученных профилей должен быть задан некоторый критерий (например, энтропия [10]), глобальный экстремум которого соответствует искомому углу наклона. Для «неверно повернутых» документов ПП содержат смазанные пики, тогда как при правильном довороте явно выделяются острые узкие пики. Примеры подобных ПП приведены на Рис. 3.

В работах [10, 11] авторы используют и вертикальную, и горизонтальную параллельные проекции. Для каждой из них вычисляются значения критерия, которые затем комбинируются для получения единой оценки угла-кандидата. Группа методов (г) опирается на анализ Хаф-образа изображения. При этом обычно используется нормальная параметризация (ρ, ϕ) аккумуляторного пространства ПХ, в которой каждой строке соответствует фиксированный угол ϕ , а вдоль оси ординат откладывается расстояние до прямой от начала координат. К ее преимуществам можно отнести: (1) возможность явно управлять шагом дискретизации угла, (2) простоту интерпретации и наглядность результата, (3) отсутствие вырожденного случая с «бесконечным» наклоном.



Рис. 3. Пример вертикальной проекции текстового фрагмента на ось абсцисс: слева – до нормализации, справа – после

Сумма значений в любой строке ПХ равна полной сумме значений яркости по всему изображению (так называемый радоновский инвариант), а информация о всех проекциях сосредоточена в единственной матрице $HT(I)$, которая может быть интерпретирована как цифровое изображение. Тогда, исходная задача сводится к выбору индекса строки $HT(I)$, соответствующей искомому направлению ϕ . Этот выбор, как и в случае анализа ПП, базируется на использовании критерия $f(r)$, который вычисляется для каждой строки r матрицы $HT(I)$. В качестве критерия $f(r)$ часто применяется сумма квадратов производной яркости [2; 11] (далее – SSG):

$$f(r) = SSG(r) = \sum_{c=1}^{\omega-1} (r[c] - r[c-1])^2. \quad (2)$$

В формуле (2) ω означает число столбцов $HT(I)$, а $r[c]$ – значение яркости пикселя в строке r и столбце c . Искомый индекс выражается следующим образом:

$$i_{res} = \underset{r \in \text{rows}(HT(I))}{\operatorname{argmax}} f(r). \quad (3)$$

Приведенной схеме следует множество методов. Типовой в этом плане является, например, работа [12]. Различия в схеме, как правило, связаны со следующими аспектами. Во-первых, часто для повышения устойчивости метода перед вычислением ПХ тем или иным образом осуществляется фильтрация исходного растра. Во-вторых, вычислительная сложность «классического» ПХ зачастую является препятствием для его применения на практике и вынуждает искать пути его более эффективного подсчета. Например, в работе [13] для уменьшения объема необходимых вычислений предложено использовать рандомизированное ПХ. Исследование [14] опирается на модификацию, известную под названием иерархическое ПХ. В работе [15] предлагается вычислять ПХ для двух диапазонов: отдельно для вертикальных углов и отдельно для горизонтальных, а затем запускать итеративный процесс уточнения угла наклона.

Серьезным недостатком большинства указанных методов является требование подавать на вход бинарный образ исходного изображения. Во-первых, эта процедура зачастую приводит к существенной потере информации, хранящейся в растре, а во-вторых, выбор и настройка

метода бинаризации является весьма нетривиальной задачей, зависящей от многих факторов. К сожалению, на сегодняшний день не существует универсального алгоритма бинаризации, хорошо подходящего для всех типов документов [1].

Для задачи D_{slant} также было разработано множество методов. Среди существующих подходов можно выделить две основные группы. Первая группа базируется на анализе компонент, выделенных каким-либо образом, внутри текстового фрагмента. Данные методы, как и в предыдущей задаче, опираются на бинаризацию исходного образа. Выделенные компоненты обычно фильтруются для того, чтобы оставить только преимущественно вертикальные элементы символов [16]. Угол наклона всего фрагмента вычисляется путем комбинирования углов наклона выделенных компонент. Такие подходы позволяют учитывать неравномерность распределения наклона в рамках одной строки [17]. К этому же классу можно отнести и метод, предложенный в работе [18], где используется анализ вейвлет разложения исходного образа. Некоторые методы, как например [19], активно опираются на особенности начертания символов в определенных языках, что делает неприменимым (или требует существенной переработки) их использование в системах распознавания с произвольными входными алфавитами.

Ко второй группе относятся подходы, совпадающие с описанной ранее схемой перебора множества углов, построении ПП и оптимизации вычисляемого по ним функционала [20]. При этом, специфика задачи позволяет сильно ограничить диапазон рассматриваемых углов (обычно это диапазон от 0° до 45°). В целях оптимизации часто обрабатывается лишь заранее выделенное подмножество углов, как правило с фиксированным шагом [19].

Применение ПХ в определении наклонного начертания текстового фрагмента гораздо менее распространено, но иногда все же встречается, например, в работе, связанной с нормализацией арабицы [21]. Однако все перечисленные работы вычисляют ПХ не от изображения в градациях серого, а от бинарной карты границ или каким-либо образом бинаризованного растра, что зачастую приводит к потере информации и

существенно усложняет схему предобработки фрагмента.

Таким образом, интерес представляет построение метода, позволяющего осуществлять вычисление проекций изображения в фиксированном диапазоне углов более эффективно, чем классическое ПХ. Серьезным плюсом будет возможность отказа от бинаризации исходного образа, а также минимальное количество настроечных параметров у алгоритма.

Предлагаемый метод

В качестве метода решения рассматриваемых задач автоматического определения угла наклона предлагается использовать анализ результата быстрой аппроксимации ДПР, вычисленного с помощью алгоритма Брейди-Ёна [5], то есть – БПХ. К сожалению, она не может быть эффективно вычислена сразу для полного диапазона углов. Так, в задаче D_{skew} , чтобы построить преобразование Хафа, т.е. посчитать проекции изображения для всех направлений, может потребоваться применить данный алгоритм четырежды (по разу на квадрант) и объединить результаты в единый Хаф-образ. Детали построения такого объединения описаны в работе [22]. На Рис. 4 справа приведен пример полного БПХ-образа. Здесь каждое из множеств Π соответствует множеству прямых, укладывающихся в четыре угловых диапазона: $[0^\circ, 45^\circ]$ для

Π_H^+ , $[45^\circ, 90^\circ]$ для Π_V^+ , $[90^\circ, 135^\circ]$ для Π_V^- и $[135^\circ, 180^\circ]$ для Π_H^- .

В задаче D_{skew} для автоматического определения угла наклона предлагается анализировать не исходное изображение, а его производную по вертикальным и горизонтальным направлениям (из соображений симметрии). Это приводит к необходимости вычисления двух различных БПХ-образов F_V и F_H по диапазонам Π_V и Π_H соответственно. В тоже время, в задаче D_{slant} более важными являются преимущественно вертикальные компоненты текстовых фрагментов, поэтому можно обойтись анализом производной, взятой только по оси x . Это позволяет нивелировать влияние преимущественно горизонтальных компонент. В ряде случаев, например, при обработке печатных фрагментов или фрагментов, написанных кириллицей, в задаче D_{slant} почти всегда встречаются наклоны только «вправо», при этом диапазон допустимых углов очевидно оказывается ограничен классом «преимущественно вертикальных» прямых, что может позволить обойтись вычислением преобразования только для единственного квадранта и тем самым существенно сократить количество операций. На Рис. 5 приведен пример результата быстрой аппроксимации ДПР, построенной с помощью алгоритма Брейди-Ёна, для производной, взятой по оси x . Квадрант соответствует прямым с наклоном «вправо».

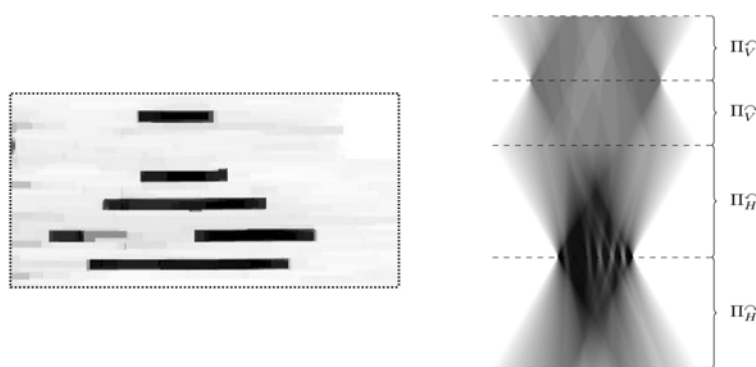


Рис. 4. Пример БПХ-образа: слева – исходное изображение; справа – полный БПХ-образ



Рис. 5. Пример вычисления БПХ-образа для одного квадранта Π_V^+

Важной особенностью используемой аппроксимации ДПР является неравномерность системы координат для результирующего аккумуляторного пространства [5]. Строке БПХ с фиксированным наклоном t в ДПР соответствует угол $\phi_t = \arctg(t/(h-1))$, где h – высота исходного изображения. Это приводит к тому, что каждая строка БПХ масштабирована относительно соответствующей строки в ДПР в $k_t = 1/\cos \phi_t$ раз. Поэтому, при вычислении критерия по отдельной строке БПХ это обстоятельство требуется учитывать и умножать исходное значение критерия на функцию от k_t , которая, в свою очередь, зависит от выбранного критерия. Для часто используемых критериев подобные функции k_t приведены в работе [23]. В частности, для критерия SSG эта функция равна k_t^3 , таким образом $SSG_{DRT}(t) \approx k_t^3 \cdot SSG_{FHT}(t)$.

Стоит отметить, что рукописное заполнение документов как правило отличается весьма высоким качеством и, в основном, имеет одно доминирующее направление, однако угол наклона может не попадать в один квадрант. Пример такого случая иллюстрирует Рис. 6. Для того, чтобы корректно обрабатывать подобные случаи предлагается использовать следующий подход: сжать исходное изображение до вычисления БПХ в два раза по горизонтали. Если до сжатия максимальный рассматриваемый угол составлял $\arctan(1) = \pi/4$, то после сжатия он увеличился до $\arctan(2)$, что соответствует примерно $63,5^\circ$ (большие углы, как правило, уже нет смысла рассматривать). Для возврата к наклону в исходном изображении достаточно просто умножить найденный на сжатой картинке наклон обратно в два раза.

Как видно из Рис. 4, для не квадратного изображения в БПХ-образе $|\Pi_H^\wedge| = |\Pi_H^\wedge|$, $|\Pi_V^\wedge| = |\Pi_V^\wedge|$, но $|\Pi_V^\wedge| \neq |\Pi_H^\wedge|$. Согласно уравнению (4) это означает, что множества углов Ω_H и Ω_V , покрываемые по горизонтали и вертикали

соответственно, могут существенно различаться. Если возникает необходимость привести их к одному диапазону, то требуется зафиксировать одно из множеств, например Ω_H , в качестве опорного, и провести соответствующую интерполяцию значений другого множества к опорному (процедура *Recalculate*). Такая необходимость может возникнуть из следующего изображения: вертикальные и горизонтальные элементы изображения зачастую перпендикулярны друг другу (линии разграфки, штрихи в печатных буквах и т.д.), поэтому следует анализировать их совместный вклад в пару перпендикулярных направлений. Совместный вклад определяется суммой критериев с предопределенными весами β и $1 - \beta$ (процедура *Combine*).

Приведенным выше описаниям соответствуют базовые алгоритмы A_{skew}^{FHT} и A_{slant}^{FHT} , предназначенные для решения задач D_{skew} и D_{slant} соответственно. Они представлены листингами на Рис. 7, из которых следует, что эти алгоритмы оказываются крайне близки.

Способы оценки точности определения угла наклона

В 2013 году был проведен конкурс Document Image Skew Estimation Contest (далее – DISEC) [24], посвященный задаче D_{skew} . Организаторы отобрали 155 различных документов, каждый из которых представлен десятью изображениями со случайными углами наклона в диапазоне $\pm 15^\circ$ с указанием соответствующего эталонного значения угла. Эти документы существенно различаются по размерам бумажного листа, содержат различные виды алфавитов и начертаний символов, на части документов присутствуют разнообразные графические элементы, а также табличные структуры. По завершению конкурса были опубликованы отранжированные результаты всех присланных методов, а сам корпус данных выложен в открытый доступ и де-факто стал базовым для задачи D_{skew} .

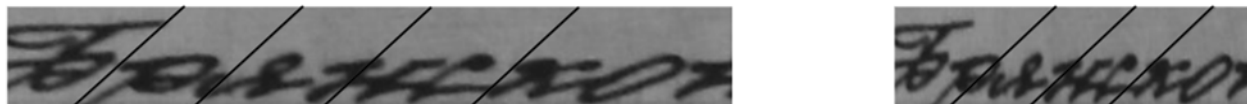


Рис. 6. Пример рукописного фрагмента с большим значением угла наклона (слева) и результат его сжатия в два раза по горизонтали (справа)

```

1 def deslanter(I, k):
2     Idown ← Downscale(I, k)
3     Dx ← HorizontalDerivative(Idown)
4
5     F⊥ ← FastHoughTransform(Dx, ΠV)
6
7     Nv ← height(F⊥)
8
9     for i in rows(F⊥):
10        KV[i] ← √(1 + i2/(Nv - 1)2)
11        CV[i] ← KV3[i] · SSGFHT(F⊥[i])
12
13    sbest ← argmax(CV)
14    sbest ← k · sbest
15    α ← arctan(sbest)
16    return α
    
```

Листинг 1: Алгоритм $\mathcal{A}_{slant}^{FHT}$ определения угла наклона текстового фрагмента с помощью БПХ.

```

1 def deskewer(I, β):
2
3     Dx ← HorizontalDerivative(I)
4     Dy ← VerticalDerivative(I)
5     F⊥ ← FastHoughTransform(Dx, ΠV)
6     F⊗ ← FastHoughTransform(Dy, ΠH)
7     Nv ← height(F⊥)
8     Nh ← height(F⊗)
9     for i in rows(F⊥):
10        KV[i] ← √(1 + i2/(Nv - 1)2)
11        CV[i] ← KV3[i] · SSGFHT(F⊥[i])
12
13    for i in rows(F⊗):
14        KH[i] ← √(1 + i2/(Nh - 1)2)
15        CH[i] ← KH3[i] · SSGFHT(F⊗[i])
16    CV ← Recalculate(CV)
17    C ← Combine(CH, β, CV, 1 - β)
18    sbest ← argmax(C)
19
20    α ← arctan(sbest)
21    return α
    
```

Листинг 2: Алгоритм \mathcal{A}_{skew}^{FHT} определения угла наклона документа с помощью БПХ.

Рис. 7. Листинги предлагаемых алгоритмов \mathcal{A}_{skew}^{FHT} и $\mathcal{A}_{slant}^{FHT}$

В рамках DISEC было предложено три критерия точности определения угла наклона [24]. Для каждого изображения вычисляется E – модуль отклонения полученного угла наклона от эталонного значения, выраженного в градусах. На множестве всех полученных модулей отклонений вычисляются следующие величины: (а) AED – среднее значение отклонения; (б) $TOP80$ – среднее значение отклонения среди 80% лучших значений; (в) CE_t – доля отклонений, меньших порога t (по утверждению организаторов порог различения ошибки в задаче D_{skew} составляет 0.1°).

$$AED = \frac{1}{N} \sum_{i=0}^N E_i;$$

$$TOP80 = \frac{1}{0.8 \cdot N} \sum_{i=0}^{0.8 \cdot N} sorted(E)_i;$$

$$CE_t = \frac{1}{N} \sum_{i=0}^N [E_i \leq t]$$

В критерии CE_t [...] означает скобку Айверсена. Значения первых двух критериев требуется минимизировать, последнего – максимизировать. В данной работе при проведении обоих вычислительных экспериментов используются эти же критерии.

Нормализация текстовых фрагментов

В 2019 году была опубликована работа С. Бера и его соавторов [25], посвященная проблеме геометрической нормализации рукописных текстовых фрагментов, в рамках которой было представлено несколько корпусов данных для проведения вычислительных экспериментов. В трех корпусах K_{BEN} , K_{HIN} и K_{ENG} содержатся изображения реальных рукописных фрагментов на бенгали, хинди и латинице соответственно. Каждый из них состоит из 500 образцов, повернутых на некоторый угол α_{skew} и скошенных на некоторый угол α_{slant} . Точные значения этих углов закодированы в имени соответствующих файлов и выступают в качестве эталона. В рамках данного эксперимента для задачи D_{slant} интерес представляет только определение угла α_{slant} . Поэтому все изображения были предварительно повернуты на эталонное значение угла, предоставленного авторами, в результате чего были сформированы корпуса K_{BEN}^{desk} , K_{HIN}^{desk} , K_{ENG}^{desk} на которых в дальнейшем и производились измерения. На Рис. 8 приведены примеры текстовых фрагментов из исходных и повернутых на эталонный угол корпусов данных.

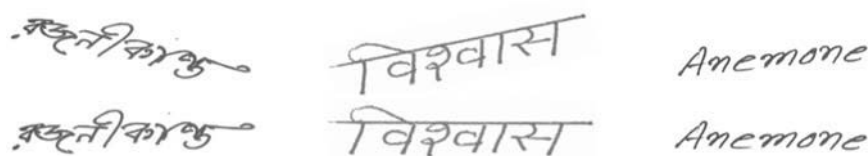


Рис. 8. Примеры изображений из трех корпусов данных, взятых из работы [25] (верхний ряд) и соответствующие результаты предварительной нормализации (нижний ряд)

К сожалению, в описанных выше корпусах данных отсутствуют кириллические текстовые фрагменты. Поэтому автором данной работы был сформирован дополнительный корпус данных K_{RUS} . Он состоит из 2750 рукописных фрагментов, извлеченных с изображений паспортов гражданина Российской Федерации, полученных с помощью сканирования и фотографирования носителей камерами мобильных устройств. Примеры текстовых фрагментов приведены ранее на Рис. 2. Все фрагменты в K_{RUS} состоят исключительно из кириллического или цифрового заполнения, для каждого фрагмента также представлен эталонный угол его наклона. Для рукописного текста, в отличие от печатных образцов, такой угол не всегда может быть указан однозначным образом ввиду неравномерности наклона внутри строки или же специфики самого содержимого. В таких случаях угол наклона указывался в соответствии с ожидаемым экспертом значением. В соответствии с распространенной

конвенцией [24; 25], значение эталонного угла закодировано в именах файлов. Все используемые в работе корпуса данных, включая предварительно нормализованные корпуса из работы [25], выложены в открытый доступ по адресу <ftp://smartengines.com/text-norm-fht>.

Распределения углов наклона для рассматриваемых в работе корпусов данных проиллюстрированы Рис. 9. Из данных рисунков видно, что распределения являются существенно различными и для достижения максимальной точности или быстрого действия детектора угла наклона могут потребоваться различные настройки для предлагаемых методов.

Для замеров точности и скорости решения задачи D_{slant} использовалось два алгоритма: A_{slant}^{FHT} и A_{slant}^{DRT} . Первый алгоритм описан листингом 1, а второй выступает в качестве референтного. По сути, он является модификацией A_{slant}^{FHT} с заменой вычисления БПХ на ДПР. При вычислении ДПР требуется явно указывать шаг угла

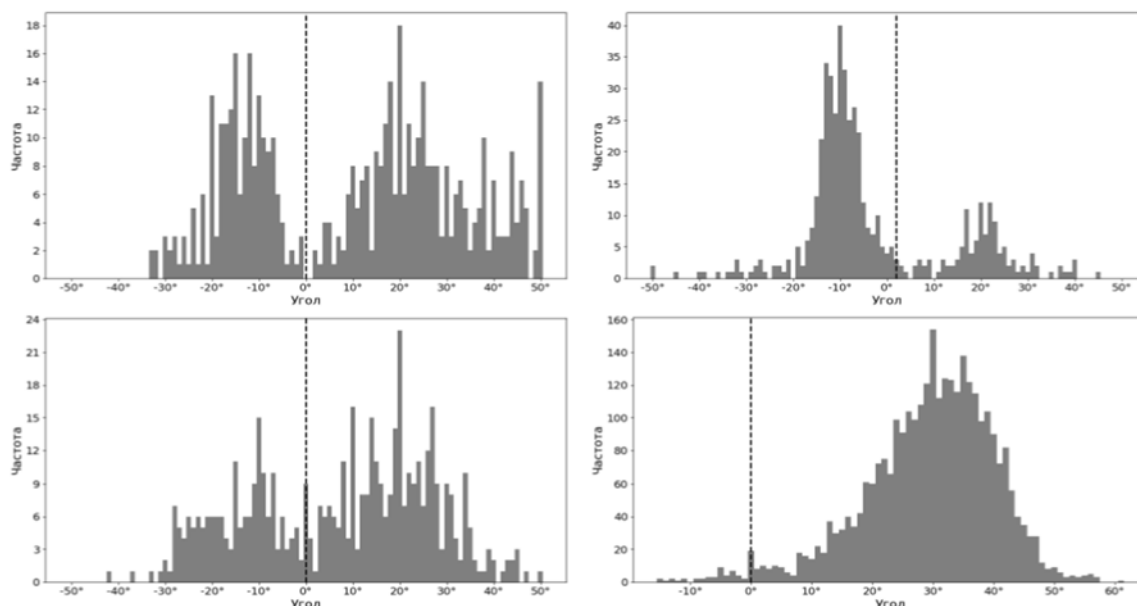


Рис. 9. Распределение угла наклона в корпусах данных: сверху слева - K_{BEN} ; сверху справа - K_{HIN} ; внизу слева - K_{ENG} ; внизу справа - K_{RUS}

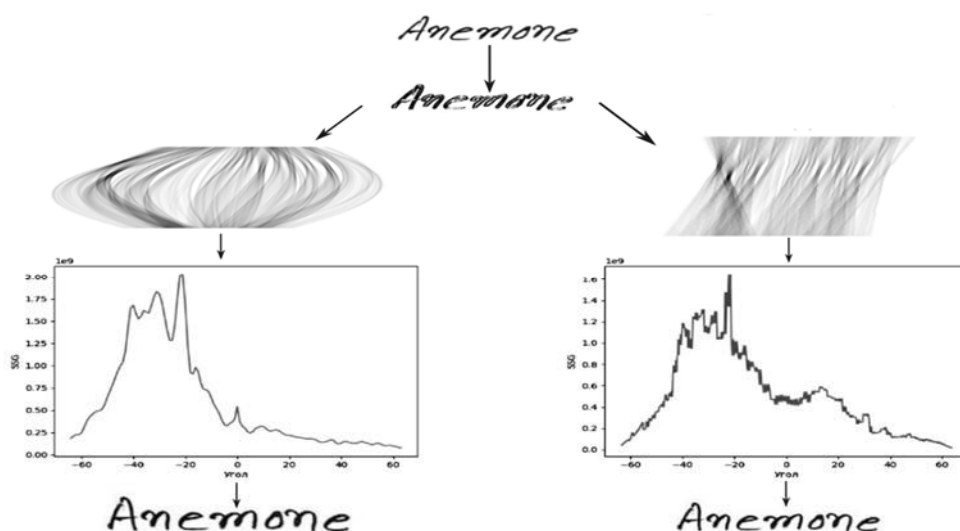


Рис. 10. Основные этапы работы алгоритмов A^{DRT}_{slant} (слева) и A^{FHT}_{slant} (справа)

(в градусах), а также минимальное и максимальное значение обрабатываемых углов. Поскольку эталон для изображений в используемых корпусах данных установлен в 1° , то и для критерия CE был установлен такой же порог различия, как и шаг угла при вычислении ДПР.

Напомним, что по построению БПХ, покрываемый им набор углов зависит от размеров входного изображения. Поскольку из изображений на Рис. 9 видно, что диапазона $[-45^\circ; +45^\circ]$ не хватает для полноценной обработки данных, то в случае A^{FHT}_{slant} используется описанный ранее прием со сжатием изображения в два раза по горизонтали, чтобы полностью покрыть весь требуемый диапазон углов (хотя из вида распределения на Рис. 9, необходимость такого шага уже не выглядит очевидной). Поэтому диапазон обрабатываемых углов для A^{DRT}_{slant} установлен в $[-64^\circ; +64^\circ]$. На Рис. 10 проиллюстрированы основные этапы процесса обработки изображения *WENG_001* со словом «Anemone». На данном примере ответы алгоритмов A^{DRT}_{slant} и A^{FHT}_{slant} совпали и равны -22° градуса, в то время как значение эталона -28° , то есть ошибка методов составила 6° .

Предлагаемые методы реализованы на языке программирования C++ с использованием открытого компилятора GCC версии 11.4.0 и открытой библиотеки компьютерного зрения OpenCV. БПХ-образы для прямых вычислены с помощью OpenCV-contrib – ее открытого расширения. Все показатели производительности получены на ПК под управлением ОС Ubuntu 22.04.4 LTS, на базе

процессора AMD Ryzen 7 2700X (8 ядер) и 32Гб оперативной памяти.

Полученные результаты работы методов нормализации наклона представлены далее в Табл. 1. Из таблицы видно, что метод на основе БПХ работает гораздо быстрее аналога с использованием ДПР (в некоторых случаях – на порядок), при этом разность в точности детекции оказывается не столь значительной. Распределения модуля отклонений угла наклона текстовых фрагментов от эталонных значений E для алгоритма A^{FHT}_{slant} для рассматриваемых корпусов данных проиллюстрированы Рис. 11. Следует отметить, что с помощью предварительной обработки изображения можно повысить точность работы детектора угла, однако делать это нужно всегда с оглядкой на особенности решаемой задачи и входных данных.

Нормализация угла наклона документа

Для вычислительных экспериментов, связанных с задачей D_{skew} , будем пользоваться описанным ранее корпусом данных DISEC. Как и в предыдущем разделе, рассмотрим как на целевом корпусе данных работают алгоритмы A^{DRT}_{skew} и A^{FHT}_{skew} , основанные на вычислении ДПР и БПХ, соответственно. При этом, учтем, что максимальное значение угла наклона в данном корпусе данных по модулю не превышает 15° , а ожидаемая точность определения угла наклона составляет 0.1° , что соответствует метрике $CE_{0.1}$.

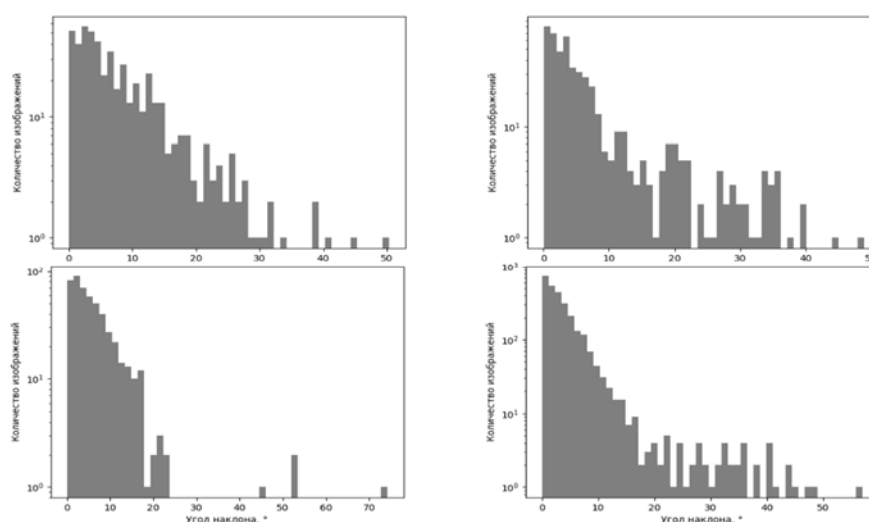


Рис. 11. Распределение модуля отклонений угла наклона текстовых фрагментов от эталонных значений E для алгоритма A_{slant}^{FHT} в корпусах данных: сверху слева - K_{BEN} ; сверху справа - K_{HIN} ; внизу слева - K_{ENG} ; внизу справа - K_{RUS}

Табл. 1. Результаты замеров на тестовых корпусах данных для задачи D_{slant} .

Алгоритм	Корпус	AED, °	TOP80, °	$CE_{1.0}$, %	t_{mean} , мс	t_{std} , мс
A_{slant}^{DRT}	K_{BEN}^{desk}	7.142	4.365	17.8	85.629	13.703
	K_{HIN}^{desk}	6.716	2.940	25.2	86.351	17.243
	K_{ENG}^{desk}	5.606	3.600	19.2	91.335	12.191
	K_{RUS}	4.729	2.082	29.7	95.673	30.915
A_{slant}^{FHT}	K_{BEN}^{desk}	7.923	4.884	10.2	2.816	1.390
	K_{HIN}^{desk}	7.010	3.336	16.0	2.704	1.269
	K_{ENG}^{desk}	6.177	3.907	11.4	4.380	2.324
	K_{RUS}	3.850	2.131	23.4	1.881	0.852

Табл. 2. Время работы метода A_{DRT}^{skew} в секундах с шагом в 1° на корпусе данных DISEC

t_{mean}	t_{std}	t_{min}	$t_{25\%}$	$t_{50\%}$	$t_{75\%}$	t_{max}
0.2243	1.1309	0.0517	0.0753	0.1064	0.1223	16.0335

Следует отметить, что по сравнению с задачей D_{slant} , линейные размеры изображений в данной задаче существенно больше. Из-за этого скорость работы методов может стать серьезной проблемой. Поэтому для начала проведем замер метода A_{skew}^{DRT} в диапазоне углов $[-15^\circ, +15^\circ]$ с шагом в 1° . В таком случае, для каждого изображения потребуется вычислить $(2 \cdot 15 - 1)$ проекций. В Табл. 2 приведено время работы данного метода на корпусе данных DISEC. Из таблицы видно, что время работы метода является неприемлемо долгим. Результаты по точности детекции угла приведены в Табл. 3. При дальнейшем

повышении углового разрешения время работы ощутимо увеличивается. Так, для изображения с размерами 1095×894 пикселей, взятого из корпуса данных DISEC, число проекций для БПХ составит 3975 (для горизонтальных и вертикальных направлений) и время его вычисления составляет 45мс, в то время как для ДПР для того же числа проекций потребуется уже 21с (данный замер был получен на ПК под управлением ОС Ubuntu версии 18.04, процессор: AMD Ryzen 7 1700 (8 ядер), 16Гб оперативной памяти). В таблице 3 приведена точность работы некоторых методов на корпусе данных DISEC.

Табл. 3. Результаты замеров некоторых методов на корпусе данных DISEC

Method	AED, °	TOP80, °	CE, %
LRDE-EPITA-a	0.072	0.046	77.48
Ajou-SNU	0.085	0.051	71.23
LRDE-EPITA-b	0.097	0.053	68.32
Gamera	0.184	0.057	68.90
CVL-TUWIEN	0.103	0.058	65.42
A_{skew}^{DRT}	0.411	0.240	18.80
A_{skew}^{FHT}	0.083	0.054	68.80

Метод LRDE-EPITA-a основан на специальной предобработке изображения и последующего анализа его Фурье-образа. Метод Ajou-SNU находит на изображении линии специального вида и оценивает угол наклона путем взвешенного голосования на множестве данных линий. Метод LRDE-EPITA-b также основан на предобработке изображения и последующим анализом образа «классического» преобразования Хафа.

Максимальная ошибка на корпусе данных составляет всего 0.547° , что указывает на то, что метод не порождает на нем грубых ошибок. В методе A_{skew}^{FHT} для БПХ вычисляется весь покрываемый диапазон углов в $[-45^\circ, +45^\circ]$, однако оце-

ниваются только те углы, что входят в допустимый диапазон по условиям задачи. Распределение ошибок E для алгоритма A_{skew}^{FHT} на корпусе данных DISEC проиллюстрировано Рис. 12.

Поскольку в DISEC каждое изображение представлено 10 различными наклонами, была произведена группировка изображений и оценивались критерии внутри получившихся групп. Пять худших групп были отобраны и представлены в Табл. 4.

Дополнительно была проведена группировка всех примеров изображений с шагом в один градус. Внутри полученных групп было оценен показатель AED. Результат приведен на Рис. 13.

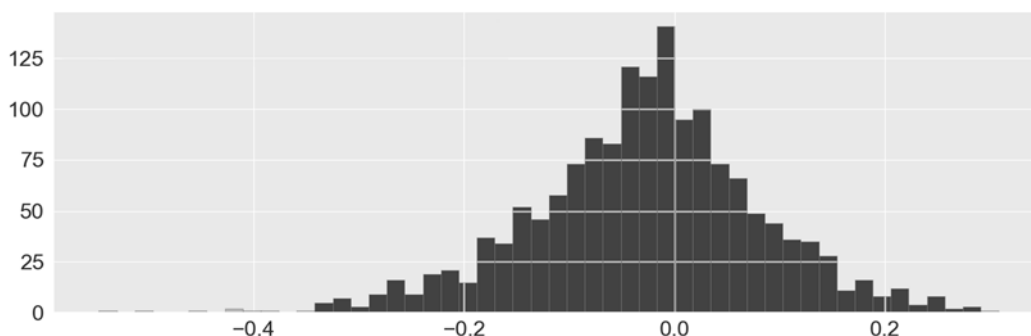


Рис. 12. Распределение ошибок E на корпусе данных DISEC

Табл. 4. Распределение ошибок по группам изображений на корпусе данных DISEC

Индекс	AED, °	TOP80, °	CE, %	$E_{MAX}, ^\circ$	$E_{MIN}, ^\circ$	$E_{RANGE}, ^\circ$
68	0.307	0.299	0.0	0.340	0.266	0.074
74	0.279	0.249	0.0	0.417	0.125	0.292
18	0.260	0.235	0.0	0.406	0.124	0.282
61	0.256	0.245	0.0	0.315	0.211	0.104
84	0.253	0.244	0.0	0.305	0.202	0.103

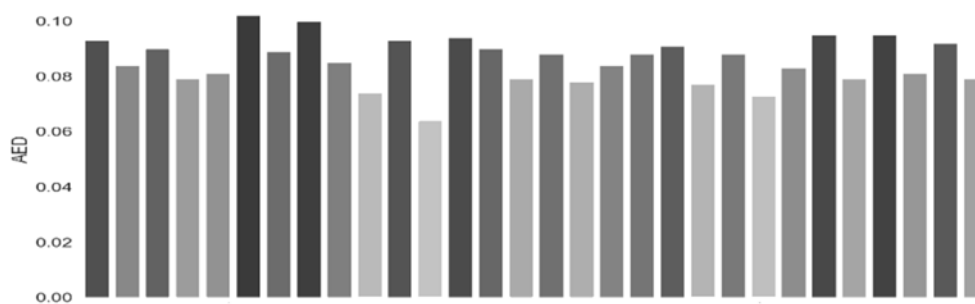


Рис. 13. Значение величины AED для каждого уровня с шагом в один градус на корпусе DISEC

Заключение

В данной работе рассмотрены способы нормализации изображений текста с использованием быстрого преобразования Хафа. Были предложены решения двух задач из области автоматической обработки изображений документов: алгоритм определения угла наклона документа и алгоритм определения угла наклона текстового фрагмента. Предложенные алгоритмы не требуют бинаризации входного изображения, применимы в широком диапазоне углов и обладают хорошим быстродействием, что позволяет использовать их в том числе и на мобильных устройствах.

Реализации предложенных методов были апробированы на открытых корпусах данных, специфичных для указанных задач. Конкретнее, алгоритм детектирования угла наклона документа апробировался на известном корпусе данных DISEC, а алгоритм определения наклона текстового фрагмента – на корпусах данных из работы [25]. Кроме того, был подготовлен и опубликован специальный кириллический корпус данных K_{RUS} , состоящий из 2750 изображений. Используемые в работе корпуса данных доступны по адресу <ftp://smartengines.com/text-norm-fht>.

Предложенные методы могут быть внедрены, например, в промышленные системы распознавания документов, удостоверяющих личность, функционирующих на мобильных устройствах, где вопросам быстродействия традиционно уделяется повышенное внимание.

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 // Компьютерная оптика. — 2022. — Т. 46. — № 4. — С. 567-589. — 10.18287/2412-6179-CO-1020.
2. Брухтий А.В., Куратов П.А. “Использование полутонного изображения при поиске границ символов”, Труды ИСА РАН, 2014, Т. 64, № 4, С. 80-83.
3. Чернов Т.С., Ильин Д.А., Безматерных П.В., Фараджев И.А., Карпенко С.М. “Исследование методов сегментации изображений текстовых блоков документов с помощью алгоритмов структурного анализа и машинного обучения”, Вестник РФФИ, 2016, № 4, С. 55-71, 10.22204/2410-4639-2016-092-04-55-71.
4. Chernyshova Y.S., Sheshkus A.V., Arlazarov V.V. Two-step CNN framework for text line recognition in camera-captured images // IEEE Access. — 2020. — Т. 8. — С. 32587-32600. — DOI: 10.1109/ACCESS.2020.2974051.
5. Brady M.L., Yong W., “Fast parallel discrete approximation algorithms for the Radon transform,” SPAA’92, 91–99, ACM, New York, NY, USA (1992).
6. Nikolaev D. P., Karpenko S. M., Nikolaev I. P., Nikolayev P. P. Hough transform: underestimated tool in the computer vision field // ECMS 2008, 2008, pp. 238-243, DOI: 10.7148 / 2008.
7. Ершов Е.И., Терехин А.П., Карпенко С.М., Николаев Д.П. “О точной оценке неточностей аппроксимации прямых в алгоритме быстрого преобразования Хафа”, ИТис 2015, 2015, ISBN 978-59-01158-28-9, С. 858-868.
8. Hull J.J., “Document image skew detection: Survey and annotated bibliography,” DAS 1996, Malvern, PA, USA, October 14-16, 1996, Selected papers., 40–66 (1996).
9. Eynard L., Emptoz H., “Italic or roman: Word style recognition without a priori knowledge for old printed documents.” in ICDAR, 823–827, IEEE Computer Society (2009).
10. Brodic D., Milivojevic Z., “Text skew detection using combined entropy algorithm,” Information Technology And Control 46 (Sept. 2017).
11. Papandreou A., Gatos B., Perantonis S., Gerardis I., “Efficient skew detection of printed document images based on novel combination of enhanced profiles,” IJdar 17, 433–454 (Sept. 2014).
12. Srihari S., Govindaraju V., “Analysis of textual images using the hough transform,” Machine Vision and Applications 2, 141–153 (jun 1989).
13. Boukharouba A., “A new algorithm for skew correction and baseline detection based on the randomized Hough

- transform,” Journal of King Saud University - Computer and Information Sciences 29, 29–38 (01 2017).
14. Yu B., Jain A., “A robust and fast skew detection algorithm for generic documents,” Pattern Recognition 29, 1599–1629 (10 1996).
 15. Stahlberg F., Vogel S., “Document skew detection based on Hough space derivatives,” in 2015 13th International Conference on Document Analysis and Recognition (ICDAR), IEEE (08 2015).
 16. Bozinovic R., Srihari S., “Off-line cursive script word recognition,” IEEE Trans. Pattern Anal. Mach. Intell. 11, 68–83 (Jan. 1989).
 17. Uchida S., Taira E., Sakoe H., “Nonuniform slant correction using dynamic programming,” in ICDAR, 434–438, IEEE Computer Society (2001).
 18. Zhang L., Lu Y., Tan C., “Italic font recognition using stroke pattern analysis on wavelet decomposed word images,” in ICPR (4), 835–838, IEEE Computer Society (2004).
 19. Fan K., Huang C., “Italic detection and rectification,” J. Inf. Sci. Eng. 23(2), 403–419 (2007).
 20. Pastor M., Toselli A., Vidal E., “Projection profile based algorithm for slant removal,” in ICIAR(2), Lecture Notes in Computer Science 3212, 183–190, Springer (2004).
 21. Stahlberg F., Vogel S., “The QCRI recognition system for handwritten arabic,” 276–286 (2015).
 22. Алиев М.А., Кунина И.А., Николаев Д.П., Полевой Д.В. “О практических аспектах вычисления Хаф-образа алгоритмом Брейди-Ена” // Информационные процессы, 2023, Т. 23, № 2, С. 250-273, 10.53921/18195822_2023_23_2_250.
 23. Бочаров Д.А. “Метод линейной регрессии, устойчивый к экстремальным стационарным помехам”, Сенсорные системы, 2020, Т. 34, № 1, С. 44-56, 10.31857/S0235009220010059.
 24. Papandreou A., Gatos B., Louloudis G., Stamatopoulos N., “ICDAR 2013 document image skew estimation contest (DISEC 2013),” in ICDAR, (2013).
 25. Bera S., Chakrabarti A., Lahiri S., Barney Smith E., Sarkar R., “Normalization of unconstrained handwritten words in terms of slope and slant correction,” Pattern Recognition Letters 128, 488–495 (Dec. 2019).

Безматерных Павел Владимирович. Федеральный исследовательский центр «Информатика и управление» РАН, Москва, Россия. Научный сотрудник-программист. Область научных интересов: анализ и обработка изображений документов, распознавание штрихкодов. E-mail: bezmpavel@gmail.com

Text Image Normalization Using Fast Hough Transform

P. V. Bezmaternykh^{1,II}

^IFederal Research Center «Computer Science and Control» of Russian Academy of Sciences, Moscow, Russia

^{II}Smart Engines Service, Moscow, Russia

Abstract. The tasks of text image normalization arise simultaneously in several modules of the automatic document image recognition system. The paper presents a solution for two classical tasks of geometric normalization of a digital text image: compensation for the global document skew angle and slant elimination for its textual fragments. For both tasks, which differ in the type of geometric distortions, the solution is based on a single method of image analysis of the fast Hough transform. This method is specified and two algorithms for solving these problems are proposed, and they are tested: for the task of slant normalization – on a variety of both known dataset and on a specially collected and published dataset of Cyrillic fragments K_{RUS} , for the task of document skew normalization – on the popular DISEC dataset. It is shown that a distinctive feature of the proposed method is high speed with the ability to process a large range of angles, and the method itself can be successfully applied in systems for automatic processing of document images.

Keywords: image normalization, fast Hough transform, document image analysis.

DOI 10.14357/20718632240401 **EDN** VOYYXQ

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, vol. 46, no 4, pp. 567-589, 2022, DOI: 10.18287/2412-6179-CO-1020.

2. A. V. Brukhtiy and P. A. Kuratov, "The method of a printed string segmentation on a grey image," *Trudy ISA RAN* 64(4), 80–83 (2014).
3. Chernov T.S., Ilin D.A., Bezmaternykh P.V., Faradzhev I.A., Karpenko S.M., "Research of Segmentation Methods for Images of Document Textual Blocks Based on the Structural Analysis and Machine Learning," *Vestnik RFFI*, № 4, pp. 55–71, 2016, DOI: 10.22204/2410-4639-2016-092-04-55-71.
4. Chernyshova Y.S., Sheshkus A.V., Arlazarov V.V. Two-step CNN framework for text line recognition in camera-captured images // *IEEE Access*. — 2020. — T. 8. — C. 32587–32600. — DOI: 10.1109/ACCESS.2020.2974051.
5. Brady M.L., Yong W., "Fast parallel discrete approximation algorithms for the Radon transform," *SPAA'92*, 91–99, ACM, New York, NY, USA (1992).
6. Nikolaev D. P., Karpenko S. M., Nikolaev I. P., Nikolayev P. P. Hough transform: underestimated tool in the computer vision field // *ECMS 2008*, 2008, pp. 238–243, DOI: 10.7148/2008.
7. Ershov E.I., Terekhin A.P., Karpenko S.M., Nikolaev D.P., "O tochnoy otsenke netochnostey approksimatsii pryamykh v algoritme bystrogo preobrazovaniya khafa," in *Proceeding of the Information Technology and Systems - 2015*, 858–868, GSP-4, 127994 (2015).
8. Hull J.J., "Document image skew detection: Survey and annotated bibliography," *DAS 1996*, Malvern, PA, USA, October 14–16, 1996, Selected papers., 40–66 (1996).
9. Eynard L., Emptoz H., "Italic or roman: Word style recognition without a priori knowledge for old printed documents," in *ICDAR*, 823–827, IEEE Computer Society (2009).
10. Brodic D., Milivojevic Z., "Text skew detection using combined entropy algorithm," *Information Technology And Control* 46, 2017.
11. Papandreou A., Gatos B., Perantonis S., Gerardis I., "Efficient skew detection of printed document images based on novel combination of enhanced profiles," *IJDAR* 17, 433–454, 2014.
12. Srihari S., Govindaraju V., "Analysis of textual images using the hough transform," *Machine Vision and Applications* 2, pp. 141–153, 1989.
13. Boukharouba A., "A new algorithm for skew correction and baseline detection based on the randomized hough transform," *Journal of King Saud University - Computer and Information Sciences* 29, 29–38, 2017.
14. Yu B., Jain A., "A robust and fast skew detection algorithm for generic documents," *Pattern Recognition* 29, 1599–1629, 1996.
15. Stahlberg F., Vogel S., "Document skew detection based on hough space derivatives," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, 2015.
16. Bozinovic R., Srihari S., "Off-line cursive script word recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* 11, pp. 68–83, 1989.
17. Uchida S., Taira E., Sakoe H., "Nonuniform slant correction using dynamic programming," in *ICDAR*, 434–438, IEEE Computer Society (2001).
18. Zhang L., Lu Y., Tan C., "Italic font recognition using stroke pattern analysis on wavelet decomposed word images," in *ICPR (4)*, 835–838, IEEE Computer Society (2004).
19. Fan K., Huang C., "Italic detection and rectification," *J. Inf. Sci. Eng.* 23(2), 403–419 (2007).
20. Pastor M., Toselli A., Vidal E., "Projection profile based algorithm for slant removal," in *ICIAR(2)*, *Lecture Notes in Computer Science* 3212, 183–190, Springer (2004).
21. Stahlberg F., Vogel S., "The QCRI recognition system for handwritten arabic," 276–286 (2015).
22. 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* 23(2), 250–273 (2023), 10.53921/18195822_2023_23_2_250.
23. Bocharov D.A., "A linear regression method robust to extreme stationary clutter," *Sensory systems* 34(1), 44–56 (2020), 10.31857/S0235009220010059.
24. Papandreou A., Gatos B., Louloudis G., Stamatopoulos N., "ICDAR 2013 document image skew estimation contest (DISEC 2013)," in *ICDAR*, (2013).
25. Bera S., Chakrabarti A., Lahiri S., Barney Smith E., Sarkar R., "Normalization of unconstrained handwritten words in terms of slope and slant correction," *Pattern Recognition Letters* 128, pp. 488–495, 2019.

Bezmaternykh Pavel V. Federal Research Center «Computer Science and Control» of Russian Academy of Sciences, Moscow, Russia. Topics of interest: document image analysis. E-mail: bezmpavel@gmail.com