ОБЗОР АЛГОРИТМОВ ДЕТЕКТИРОВАНИЯ ТЕКСТОВЫХ ОБЛАСТЕЙ НА ИЗОБРАЖЕНИЯХ И ВИДЕОЗАПИСЯХ

Ю.А. Болотова[1], В.Г. Спицын[1], П.М. Осина[1]
[1]Томский политехнический университет, Томск, Россия

Аннотация

Статья посвящена обзору методов детектирования и сегментации текстовых областей на изображениях и видеозаписях. Определяется обобщенный алгоритм работы систем распознавания текстов. Проводится обзор методов детектирования, определения структуры и сегментации текстовых документов в рамках решения задачи распознавания текстовых областей на изображениях и видеозаписях. Методы, предложенные в течение 30 лет исследований, анализируются с точки зрения точности, скорости и универсальности. В работе затрагиваются современные проблемы, касающиеся детектирования и распознавания текстовых областей на изображениях.

Ключевые слова: распознавание образов, анализ структуры документа, сегментация текстовых изображений, определение угла наклона текста.

Цитирование: **Болотова, Ю.А.** Обзор алгоритмов детектирования текстовых областей на изображениях и видеозаписях / Ю.А. Болотова, В.Г. Спицын, П.М. Осина // Компьютерная оптика. – 2017. – Т. 41, № 3. – С. 441-452. – DOI: 10.18287/2412-6179-2017-41-3-441-452.

Введение

Широкое распространение цифровых камер, средств оцифровки и сканирования привело к активному развитию методов детектирования и распознавания объектов на изображениях. Текстовая информация по-прежнему является наиболее надежным признаком для сопоставления, классификации и описания изображений. В связи с этим с 2013 года, казалось бы, угасший интерес к методам распознавания печатных и рукописных текстов снова возрастает [1, 4 – 9], существующие методы и алгоритмы пересматриваются и улучшаются с точки зрения решения задач поиска, распознавания и выявления смысла разнообразной текстовой информации на изображениях и видеозаписях [1 – 6, 8 – 11].

В настоящее время наиболее успешные системы оптического распознавания символов (OCR-системы) распознают тексты на изображениях, составленных из стандартных шрифтов, не подвергнутых шуму и искажениям, с достаточно высокой точностью. Однако даже при точности 99,9 % анализ 1 страницы текста (1500 символов) влечет за собой, в среднем, 1-2 ошибки. Для увеличения точности OCR-системы дополняются разнообразными словарными проверками, что, однако, не приводит к 100 % точности распознавания, но может инициировать возникновение новых ошибок. Следовательно, контроль результатов распознавания человеком все еще остается необходимым.

Основная задача OCR-систем заключается в назначении фрагменту изображения текста соответствующей символьной информации. Основные требования к OCR-системам включают возможности сохранения форматирования исходных документов: шрифтов, абзацев, таблиц, графиков и изображений. Современные системы распознавания поддерживают все известные текстовые и графические форматы и электронные таблицы [7].

При первичном рассмотрении систем распознавания текстов можно выделить общий алгоритм их работы, состоящий из следующих этапов:

1. Поиск области, содержащей текст, его локализация.
2. Предварительное улучшение качества локализованной области, её бинаризация.
3. Выявление структуры найденного блока текста, определение порядка чтения.
4. Сегментация текста на строки/слова и символы.
5. Получение признакового описания каждого символа.
6. Распознавание отдельных символов.
7. Словарная проверка.

Первым этапом является поиск областей текста, в результате которого на исходном изображении выделяются текстовые области. Для решения данной задачи применяются сверточные нейронные сети [9], детектор Виолы–Джонса, методы классификации на основе анализа текстурных признаков [25, 44], признаков Тамура, HOG-дескрипторов. Стоит отметить, что современные OCR-системы плохо справляются с задачей детектирования текста на произвольных изображениях, так как изначально разрабатывались для работы с изображениями, содержащими преимущественно однородный структурированный текст, что редко встречается на фотографиях и видеозаписях реальных сцен.

Алгоритмы улучшения качества предполагают предварительное сглаживание изображения, определение типов присутствующих шумов и их устранение.

Сегментация заключается в разбиении текста на строки, слова и символы. Поиск строк, как правило, основывается на периодичности и регулярности текстовых областей и осуществляется на основе метода Хафа, метода связных компонент [13], анализа горизонтальных, вертикальных и диагональных гистограмм.

Алгоритмы получения признакового описания символов разделяются на 2 класса: анализирующие исходное изображение символа в качестве признака и анализирующие вычисляемые признаки, такие как длина хорды [11], аппроксимированные контуры, остовы символов. Для снижения размерности простран-

ства признаков часто применяется метод главных компонент или линейный дискриминантный анализ.

Для распознавания символов реализуются различные классификаторы, основанные на нейронных сетях [50] и наиболее распространенных на сегодняшний день свёрточных нейронных сетях [1, 9].

Словарная проверка осуществляется на основе стандартных или динамически созданных языковых словарей, *N*-грамм, реализованных в виде списков, деревьев или графов [10, 12].

В данной работе рассматриваются алгоритмы наилучших на сегодняшний день OCR-систем, как коммерческих: ABBYY FineReader, OmniPage, Readiris Pro7, так и открытых: CuneiForm, Ocropus, Tesseract, и анализируются заложенные в них методы, направленные на решение следующих задач: поиск и локализация текста, определение структуры текста, определение угла наклона текстовых строк.

Задача определения структуры документа, как правило, нацелена на выделение однородных блоков, таких как текст, изображение, график, таблица и т.д.

Алгоритмы классификации блоков на текст и нетекст позволяют определить, является ли выделенный блок текстом, изображением или графиком, таблицей и др.

1. Современные подходы к оптическому распознаванию символов

Общий алгоритм работы OCR-систем на произвольных изображениях приведен на рис. 1. Первым шагом является отделение объектов переднего плана от фона. Далее происходит определение типа выделенных областей. Если найденная область является текстом, то можно приступать к ее сегментации и распознаванию. В процессе своего развития подходы к OCR-системам корректировались, проходя через следующие этапы [13]:

1. Последовательное выполнение основных шагов без возможности редактирования результатов работы на предыдущих этапах.
2. Последовательное выполнение основного алгоритма с возможностью возвращения на некоторые предыдущие этапы (например, возвращение на этап сегментации при неудовлетворительных результатах распознавания).
3. Подходы без предварительной сегментации. Документ сегментируется на строки, а строки подаются на вход алгоритму распознавания без сегментации на слова и символы. В подобных случаях классификатором могут быть иерархические скрытые модели Маркова, рекуррентные сети [16], свёрточные нейронные сети [54].
4. Подходы без предварительной сегментации на слова и символы с возможностью пересмотра предыдущих решений по разделению документа на строки.

Современные подходы к оптическому распознаванию символов можно разделить на две основные группы: подходы, использующие и не использующие сегментацию на слова и символы.

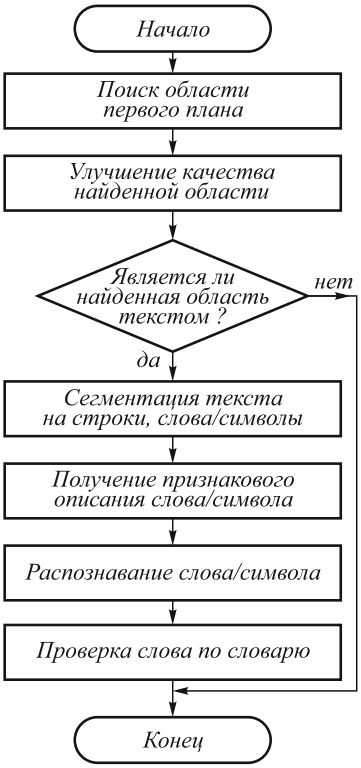

Рис. 1. Общий алгоритм работы OCR-системы

Оба подхода имеют определенные достоинства и недостатки. К наиболее значимым недостаткам подходов, использующих сегментацию, относятся возможные ошибки сегментации, влекущие за собой снижение точности последующего распознавания и дополнительные временные затраты. Обнаружение и устранение ошибок сегментации может быть сравнимым или даже более затратным по времени, чем распознавание символов. Однако в этом случае для распознавания можно использовать более простые методы, такие как анализ гистограмм, сопоставление с шаблонами, анализ различных статистических характеристик символов.

К современным системам, включающим сегментацию, относятся OCRopus [14], Tesseract [13], ABBYY FineReader, OmniPage, Readiris Pro7.

Системы без предварительной сегментации не несут временных затрат на сегментацию, но как минимум требуют определения угла наклона текста или предварительной нормализации строк, что может повлечь за собой некоторое ухудшение качества изображения. Такой подход требует применения таких моделей, как сверточные нейронные сети, скрытые модели Маркова [50], рекуррентные сети [14, 15, 47]. К достоинствам моделей данного класса можно отнести универсальность и высокую скорость работы, что в общем случае позволяет достичь более точных результатов распознавания по сравнению с моделями, требующими предварительной сегментации.

Далее будут рассмотрены методы определения объектов первого плана, классификация найденных областей на «текст» и «не текст», определения наклона текстовых строк. Результатом работы этих этапов

является представление исходного изображения, содержащего текст, в виде последовательного набора текстовых областей согласно порядку их чтения с указанием угла наклона строк.

2. Определение объектов первого плана

Методы по определению структуры документа выделяют области, содержащие объекты первого плана (текст, изображение или график), от фона на исходном изображении.

Все методы по определению объектов первого плана можно разделить на 2 категории: методы, осуществляющие анализ снизу-вверх и сверху-вниз [17]. В методах, осуществляющих анализ сверху-вниз, исходное изображение итеративно разбивается на отдельные блоки. В методах, осуществляющих анализ снизу-вверх, сначала выделяются наименьшие возможные компоненты, которые впоследствии группируются в блоки. Существуют также и гибридные подходы, комбинирующие оба этих подхода.

Методы сегментации страницы могут базироваться на анализе проекций [18–20], методе связных компонент [21], анализе пробельных прямоугольников [26], анализе контуров [21] или текстуры.

2.1. Алгоритм анализа проекций (smearing)

Одним из первых был предложен алгоритм анализа проекций. Его наиболее быстрая версия работает с RLE-представлением исходного бинарного изображения и состоит из следующих этапов [22]:

1) получение RLE-представления изображения;
2) удаление белых RLE-последовательностей, длина которых меньше порога $T1$;
3) поворот исходного изображения на 90°. Повторение шагов 1, 2 с порогом $T2$;
4) применение логической операции *"И"* над изображениями, полученными на шаге 2 и 3;
5) удаление белых RLE-последовательностей, длина которых меньше $T3$ на изображении, полученном на шаге 4 ($T3 < T1$);
6) применение метода связных компонент;
7) сопоставление площади и высоты каждой найденной области с константами для определения текстовых и нетекстовых областей.

В статье [22] авторами были подобраны следующие значения порогов $T1 = 300$, $T2 = 500$ и $T3 = 30$ для изображений с разрешением 240 dpi. Преимуществом алгоритма является его простота и высокая скорость работы за счет предварительного сжатия алгоритмом RLE. Сложность алгоритма является линейной $O(N)$. Одним из недостатков является необходимость решения вопроса о способе его бинаризации. Кроме того, если рисунок или области текста расположены достаточно близко, то высока вероятность их слияния.

2.2. Рекурсивный алгоритм X-Y cut

Рекурсивный алгоритм **X-Y cut**, впервые описанный в 1984 г. [23], относится к алгоритмам, работающим сверху-вниз. В результате его работы сегментированная страница представляется в виде дерева. На каждом этапе страница рекурсивно разбивается на две или более прямоугольные области, формирующие узлы дерева, согласно следующему алгоритму:

1) удаление мелкого шума с изображения;
2) выделение связных компонент;
3) вычисление средней высоты и ширины символа (связной компоненты);
4) поиск возможности разделения блока вертикальным или горизонтальным разрезом (возможность определяется на основе анализа построенных горизонтальных и вертикальных гистограмм блока по чёрным и белым пикселям);
5) пока есть возможность разделения, повторить шаг 4 для каждого из полученных блоков.

Наиболее сложными этапами данного алгоритма является поиск связных компонент и рекурсивное разделение блоков. Реализация двухпроходного метода связных компонент имеет сложность $O(N)$. Рекурсивное разделение на блоки аналогично построению бинарного дерева со сложностью $O(N)$. Данные этапы выполняются последовательно, следовательно, общая сложность алгоритма равна $O(N)$. Основным недостатком данного алгоритма является невозможность осуществления сегментации в случае, если ни одна из разделительных полос не разделяет страницу полностью по горизонтали или вертикали.

2.3. Алгоритм поиска максимальных белых прямоугольников

В данном алгоритме изначально предполагается наличие блочной структуры текста, в которой текстовые области разделены прямоугольными перегородками. Алгоритм состоит из двух основных этапов: поиск возможных белых прямоугольников и выбор максимальных среди них. Алгоритм поиска белых многоугольников на изображении документа впервые был предложен в [24] и адаптирован для изображений под наклоном в работе [26].

На вход алгоритма подается набор чёрных прямоугольников, содержащих связные компоненты символов документа, предварительно подвергнутого высокочастотной фильтрации. Чёрные прямоугольники сортируются по строкам и столбцам, после чего вычисляется попарное расстояние между соседними из них. Наибольший белый прямоугольник, ограниченный найденными связными компонентами, определяется по количеству связных чёрных компонент, соприкасающихся с его длинной стороной [27].

Далее из всех найденных белых прямоугольников определяются максимальные. Белый прямоугольник считается максимальным, если он не может быть больше расширен, то есть его границы соприкасаются с найденными связными компонентами или достигают края документа.

Пусть $Q(r)$ – признак качества для текущего прямоугольника r, его значение должно удовлетворять следующему условию: если $r1 \subseteq r2$, то $Q(r1) \le Q(r2)$ [28]. В качестве подобных признаков можно взять площадь, периметр или сумму квадратов высоты и ширины прямоугольника.

В работе [28] для оценки качества была предложена следующая формула (1):

$$Q(r)=\sqrt{\text{area}(r)\cdot W(\left|\log_2(\text{height}(r)/\text{width}(r)\right|)},\quad (1)$$

где W – оценка пропорциональности. В работе [29] предлагается следующая функция для оценки пропорциональности (2):

$$W(x)=\begin{cases}0{,}5, & x<3,\\ 1{,}5, & 3\le x<5,\\ 1, & \text{иначе}.\end{cases}\quad (2)$$

Таким образом, задача поиска максимального прямоугольника сводится к поиску прямоугольника, не перекрывающего никакие другие прямоугольники, функция качества которого является максимальной.

1. В очередь заносится исходное изображение.
2. Осуществляется поиск максимальных прямоугольников внутри страницы.
3. Создается очередь с приоритетами для прямоугольников (первыми сохраняются прямоугольники с наибольшим приоритетом).
4. Далее в цикле:
   4.1. Берется первый прямоугольник из очереди.
   4.2. Если он пуст (не содержит вложенных прямоугольников), то один из максимальных белых прямоугольников найден.
   4.3. Если нет, то из связных областей в исследуемом прямоугольнике выбирается один «опорный» прямоугольник, расположенный ближе к центру прямоугольника-родителя.
5. В случае пересечения опорного прямоугольника с другими ограничивается его высота и ширина.
6. Прямоугольники, с которыми он пересекается, добавляются в очередь.

Условием выхода из цикла может быть или достаточное количество найденных прямоугольников, или качество найденных прямоугольников.

В алгоритме уточнения белых прямоугольников, предложенном в [30], используется дополнительный список L:

1. Поместить все найденные максимальные белые прямоугольники в приоритетную очередь.
2. Из очереди вынимаются прямоугольники, к каждому из них применяется правило:
   2.1. Если этот прямоугольник имеет пересечения с какими-то элементами списка L, то они вычитаются из него, а результат разности передается обратно в очередь.
   2.2. Иначе текущий прямоугольник добавляется в список L.
   2.3. Пока не выполнено условие останова, перейти к Шагу 2.
3. Со страницы удаляются все прямоугольники из списка L.

Оставшиеся связные области на изображении в дальнейшем классифицируются на текстовые и нетекстовые.

К недостаткам алгоритма относится невозможность его работы с изображениями, содержащими маленькие или нерегулярные блоковые разделители, а также сложность алгоритма $O(N^2)$. Дополнительную сложность представляет поиск прямоугольников на изображениях под наклоном.

2.4. Алгоритм Dostrum

Алгоритм [31], изначально разработанный для страниц, отсканированных под небольшим наклоном, состоит из следующих этапов:

1. Удаление мелкого шума и больших явно нетекстовых объектов.
2. Применение метода связных компонент для выделения отдельных символов.
3. Кластеризация символов по размеру на 2 основные группы: символы обычного текста и заголовочные символы.
4. Поиск k ближайших соседей для каждой найденной связной компоненты.
5. Анализ пробельных символов:
   5.1. Расчет расстояния от каждой буквы до k (обычно $k=4$ или 5) ближайших соседей из того же кластера.
   5.2. На основе найденных расстояний строится гистограмма.
   5.3. К гистограмме применяется сглаживание Гаусса.
   5.4. На основе 3 первых пиков гистограмм формируются пороговые значения расстояний между буквами, строками и словами.
6. Для сегментации страницы производится транзитивное замыкание по близким компонентам (с учетом пороговых значений), в результате которого происходит объединение слов внутри строк и близких строк друг с другом.

Несомненным достоинством данного алгоритма является выбор пороговых значений согласно высотам символов. Кроме того, возможность выявления структуры изображений, находящихся под наклоном, делает данный алгоритм выигрышным по сравнению с приведенными ранее. Сложность алгоритма обусловлена наличием метода связных компонент и алгоритма к-средних и равна $O(N)$.

2.5. Диаграмма Вороного

Диаграмма Вороного представляет собой подход снизу-вверх, так как строится по центрам или крайним точкам связных компонент, называемых *опорным точкам*. Диаграмма представляет собой разбиение плоскости на области, каждая из которых соответствует одной опорной точке и является множеством точек плоскости, для которых данная опорная точка ближе, чем любая другая [32]. В статье [33] впервые был описан алгоритм применения диаграммы Вороного для решения задачи сегментации изображений, содержащих текст, состоящий из следующих этапов:

1. Удаление шума и мелких объектов.
2. Определение центров связных компонент, которые выбираются в качестве опорных точек.

3. Построение диаграммы Вороного по выбранным опорным точкам.
4. Поиск ячеек для объединения. Объединяем ячейки в двух случаях.
   4.1. Если расстояние d между опорными точками меньше порога T_1 (T_1 подбирается так, чтобы согласно этому правилу были объединены соседние буквы в словах).
   4.2. Если для d верно:

$$d/T_2 + k/T_k < 1, \quad (3)$$

где k – отношение площади большей связной области к площади меньшей (т.е. $k \geq 1$), а T_2 и T_k – пороги (если слова имеют разную высоту, то производится дополнительная проверка).

Результирующая сегментация получается в результате объединения всех ячеек, удовлетворяющих правилам 4.1 или 4.2.

Несомненным достоинством метода является то, что он содержит только три явных параметра. Исходя из проведенных экспериментов, можно сделать вывод, что диаграмма Вороного заметно опережает по качеству все вышеприведенные алгоритмы [34, 35].

Существует множество вариантов реализации данного алгоритма. Наиболее эффективным является рекурсивный алгоритм построения диаграммы Вороного, имеющий сложность $O(N*\log(N))$.

В ряде рассмотренных работ предложены некоторые способы упрощения алгоритмов, включающие снижение разрешения изображения для определения структуры документа [36], предварительное удаление шума и мелких деталей с изображения, морфологическую предобработку исходного изображения.

3. Классификация текстовых и нетекстовых областей

Выделенные на предыдущем этапе области могут представлять собой текст, горизонтальные и вертикальные линии, рисунок или график. Важным фактором при классификации областей является выбор признакового пространства. Все существующие методы классификации можно разделить на 2 категории: методы, основанные на статистических признаках; методы, основанные на спектральных признаках (преобразование Фурье, вейвлет-анализ, фильтры Габора).

Для классификации сегментированных блоков часто применяются линейные классификаторы [22], деревья решений [36] или нейронные сети [37, 50].

3.1. Алгоритмы, основанные на анализе статистических признаков

Простейший вариант алгоритма, основанный на анализе статистических признаков [27], состоит из трёх основных шагов:

1. Поиск строк на основе анализа координат найденных связных компонент.
2. Определение плотности символов в строке:

$$d = \sum w_{ch} / l_{str}, \quad (4)$$

где w_{ch} – ширина текущего символа строки, а l_{str} – длина всей строки, суммирование ведется по всем символам, входящим в текущую строку.
   2.1. Если $d > 80\%$, то это строка.
3. Если количество строк в области превышает 50%, то область классифицируется как текстовая.

В работах [22, 38] предполагается, что блок может представлять собой строку текста, горизонтальную или вертикальную линию, рисунок или график. Классификация осуществляется на основе 10 числовых признаков, рассчитанных для каждого блока:

1) высота (Δy),
2) длина (Δx),
3) площадь ($\Delta y \times \Delta x$),
4) эксцентриситет ($\Delta y/\Delta x$),
5) общее число чёрных пикселей на изображении с уменьшенным разрешением,
6) общее число чёрных пикселей в сегментированном блоке,
7) число переходов от белого к чёрному на изображении с уменьшенным разрешением,
8) процент чёрных пикселей на всем изображении с уменьшенным разрешением,
9) процент чёрных пикселей в сегментированном блоке,
10) средняя длина строк на всем изображении с уменьшенным разрешением.

Одним из основных недостатков описанного метода является невозможность правильно классифицировать области текста, содержащие более одной строки.

При дополнительном анализе текстурных признаков области можно успешно анализировать блоки любого размера (включающие несколько строк). Подход, предложенный в [39], предполагает построение двух дополнительных матриц: **BW** (оценка перехода чёрный-белый), **BWB** (оценка перехода чёрный-белый-чёрный) на основе RLE-изображения. Чёрно-белый переход (**BW**) – это набор пикселей, состоящий из множества чёрных пикселей, за которыми следует множество белых (рис. 2). Длина перехода равна суммарному количеству чёрных и белых пикселей в нем.

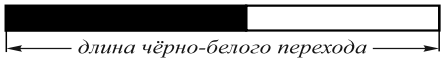

Рис. 2. Пример перехода чёрный-белый (black-white)

Выделяют 9 категорий различных переходов (согласно процентному соотношению в них чёрных и белых пикселей). Элемент матрицы **BW**(i, j) определяет количество переходов, попадающих в категорию i, длина которых равна j.

Матрица **BWB** служит для оценки распределения белых промежутков между чёрными. Длина перехода рассчитывается как длина белого промежутка в переходе. Все найденные переходы квантуются на 3 категории, исходя из процентного соотношения чёрных промежутков. Элемент матрицы **BWB**(i, j) определяет количество переходов, попадающих в категорию i, длина белой части которого равна j (рис. 3).

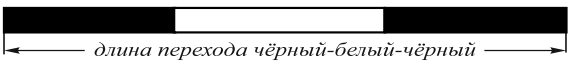

Рис. 3. Пример перехода чёрный-белый-чёрный (black-white-black)

Из таблицы **BW** рассчитываются признаки F_1 и F_2:

$$F_1 = \sum_{i=1}^{Nc}\sum_{j=1}^{Nr}\left[\frac{BW(i,j)}{j^2}\right] / \sum_{i=1}^{Nc}\sum_{j=1}^{Nr} BW(i,j), \quad (5)$$

$$F_2 = \sum_{i=1}^{Nc}\sum_{j=1}^{Nr}\left[j^2 \cdot BW(i,j)\right] / \sum_{i=1}^{Nc}\sum_{j=1}^{Nr} BW(i,j). \quad (6)$$

Из таблицы **BWB** рассчитывается признак F_3:

$$F_3 = \sum_{j=T_1}^{Nr} j^2 \left[\sum_{i=1}^{Nc} p'(i,j)\right] / \sum_{j=T_1}^{Nr}\sum_{i=1}^{Nc} p'(i,j), \quad (7)$$

где

$$p'(i,j) = \begin{cases} BWB(i,j), & BWB(i,j) > T_2, \\ 0, & BWB(i,j) \le T_2, \end{cases} \quad (8)$$

Nc – количество различных переходов (столбцов матриц **BW** и **BWB** соответственно), Nr – количество различных длин переходов (строк матриц **BW** и **BWB** соответственно).

Порог T_1 отсекает короткие переходы, т.к. наличие длинных переходов необходимо для определения графических блоков. Порог T_2 необходим для удаления небольших (случайных) значений **BWB**(i,j) длинных переходов. Экспериментально были получены значения $T_1 = 50$, $T_2 = 15$. Полученный трехмерный вектор (F_1, F_2, F_3) подается на вход линейного классификатора с целью классификации блока на 5 классов: текстовая область с маленькими, средними или большими символами, график или полутоновое изображение.

Похожий метод, заключающийся в разбиении бинарного изображения на квадратные области (10×10, 20×20 в зависимости от размера исходного изображения), был предложен в [40, 41]. В каждом окне рассчитываются следующие параметры: соотношение чёрных и белых пикселей, средняя длина чёрных промежутков, кросс-корреляция между последовательными вертикальными и горизонтальными строками, расположенными на равных промежутках друг от друга. Сложность алгоритма равна $O(N)$.

В качестве признаков можно анализировать гистограммы, оценивая долю светлых пикселей в области [42]. Метод успешно работает для отделения графиков и текста от изображений, однако документ может содержать графики с большой площадью белого фона, что создает трудность при разделении графика и текста. Дополнительно в качестве признака можно учитывать горизонтальные и вертикальные проекции блока [43].

3.2. Алгоритмы, основанные на спектральных признаках

Альтернативный подход основывается на определении спектральных признаков сегментированной области [46]. Спектральные признаки описываются через разложение изображения по набору базисов. К ним относятся косинусное разложение, синусное разложение, вейвлет-преобразование Хаара, Добеши, фильтры Габора. В процессе получения признаков выбирается одно из разложений изображения как двумерной функции, коэффициенты разложения используются в качестве признаков, характеризующих текстуру.

Первоначально классификация выполняется для каждого пикселя изображения в одну из следующих категорий: текст, рисунок, полутоновое изображение, фон, на основе анализа текстурных признаков окрестности пикселя с помощью специальных фильтров. Далее пиксели группируются согласно их пространственному расположению. Окончательное решение принимается с помощью некоторого классификатора, например, в работе [46] используется нейронная сеть. Основными недостатками данного подхода является невозможность правильно определить категорию в случае сливающихся и перекрывающихся символов, а также вычислительная сложность алгоритма.

При применении кратномасштабных ортогональных вейвлет-пакетов, подобранных согласно условию максимизации межклассовых различий, исходное изображение подвергается вейвлет-преобразованию. Вектор признаков формируется на основе расчета второго и третьего центрального момента с помощью скользящего окна в каждом диапазоне частот. Классификатор основан на нейронной сети, обученной методом обратного распространения ошибки. К преимуществам данного метода можно отнести независимость от структуры документа и хорошие результаты работы с перекрывающимися областями [45].

Классификация областей на основе фильтра Габора приведена в статье [46]. Исходное изображение подвергается фильтрации Габора с различными характеристиками (в работе рассматривается банк из 8 узкополосых фильтров), после чего для каждого полученного изображения рассчитывается локальная энергия для каждого пикселя по его окрестности. Кластеризация осуществляется для каждого отдельного пикселя исходя из значений его энергии и пространственного расположения. Результаты кластеризации достаточно точны, однако алгоритм является ресурсоемким.

Основной недостаток методов, основанных на анализе спектральных признаков, – это высокая вычислительная сложность, однако их преимуществом является достаточно точное и интуитивно понятное описание текстуры области, так как они, в первую очередь, выявляют периодичность текстуры, что является репрезентативным признаком текстовых областей, даже на изображениях с небольшим процентом текстовых областей.

3.3. Алгоритм, основанный на дискретном косинусном преобразовании

Часто цифровые изображения и видеозаписи хранятся в сжатом виде для экономии пространства и их эффективной передачи. Дискретное косинусное преобразование (ДКП) является одним из ортогональных преобразований, реализующих разложение изображения по частотам с определенными коэффициента-

ми, применяемых в алгоритмах сжатия с потерями при кодировании форматов MPEG и JPEG.

ДКП раскладывает изображение на коэффициенты, характеризующие амплитуды содержащихся в нем частот. В результате получается разреженная матрица, большинство коэффициентов которой близко или равно нулю. Ввиду того, как зрительная система человека слабо распознает определенные частоты, можно аппроксимировать некоторые коэффициенты без заметной потери качества изображения. Для этого производится квантование коэффициентов. При квантовании часть информации теряется, за счет чего достигается большая степень сжатия.

В форматах JPEG и MPEG коэффициенты ДКП рассчитываются в блоках изображения размером 8×8 [55]. Двумерное дискретное косинусное преобразование матрицы **A** размера $M \times N$ осуществляется по следующей формуле:

$$B_{pq} = \alpha_p \beta_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \times \\ \times \cos \frac{\pi (2m+1) p}{2M} \cos \frac{\pi (2n+1) q}{2N}, \quad (9)$$

где $0 \le p \le M-1$, $0 \le q \le N-1$ – найденные частоты вдоль горизонтального и вертикального направления соответственно, а

$$\alpha_p = \begin{cases} 1/\sqrt{M}, & p = 0, \\ \sqrt{2/M}, & 1 \le p \le M-1, \end{cases} \quad (10)$$

$$\beta_q = \begin{cases} 1/\sqrt{N}, & q = 0, \\ \sqrt{2/N}, & 1 \le q \le N-1. \end{cases} \quad (11)$$

Значения B_{qp}, вычисленные в результате ДКП матрицы **A**, определяют частоту и направленность блока изображения. Коэффициенты B_{qp} можно рассматривать как веса при каждой базисной функции.

В результате применения ДКП к изображению получается двумерная матрица коэффициентов, размерность которой совпадает с размерностью исходного изображения, в которой низкочастотные компоненты расположены в левом верхнем углу, а высокочастотные – справа и внизу.

Так как форматы JPEG и MPEG являются достаточно распространенными, коэффициенты ДКП целесообразно использовать в качестве признаков для поиска текстовых областей на изображениях и видеозаписях. В этом случае нет необходимости осуществлять полную декомпрессию изображения, что приводит к значительной экономии времени [54].

Для реализации ДКП требуется два вложенных цикла. Тело циклов будет выполняться $N \times N$ раз для каждого элемента матрицы дискретного косинусного преобразования. Существуют более эффективные варианты реализации ДКП, имеющие сложность $O(N^* \log(N))$.

Полученные частотные коэффициенты содержат информацию о направленности и периодичности локальных блоков изображения. Именно на этом факторе основывается возможность применения ДКП для детектирования текстовой информации, так как текстовые области обладают уникальными текстурными особенностями. Они, как правило, состоят из текстовых строк определенной ориентации с постоянным межсимвольным и междустрочным расстоянием, символы расположены последовательно друг за другом. Можно предположить, что анализ полученных коэффициентов ДКП позволит детектировать текстовую информацию на изображении более эффективно.

После первичной оценки к блокам-кандидатам на текстовую область можно применить морфологические операции, метод связных компонент, после чего подавать на вход классификатора, например на нейронную или свёрточную нейронную сеть для детектирования и распознавания текста.

К недостаткам данного метода относится невозможность детектирования текстовых символов, соразмерных с исходным изображением, так как в этом случае будет невозможно отделить периодичность и структурность текстовой области.

4. Определение угла наклона строк текста

После того, как текстовые блоки были найдены, для осуществления сегментации необходимо определить их направленность. Нормализация, представляющая собой выравнивание строк вдоль горизонтали, часто приводит к снижению качества изображения. Поэтому лучше ограничиться определением угла наклона строк текста и выполнять сегментацию с его учетом. Для решения данной задачи чаще всего используются алгоритмы анализа проекций, преобразование Хафа, метод связных компонент, оценка межстрочной корреляции [45].

Анализ горизонтальных и вертикальных проекций строк [47, 49] является одним из простейших способов обнаружения угла наклона, однако он требует построения и оценки проекций для каждого возможного угла наклона изображения, что является затратным по времени.

Более успешно для решения этой задачи применяется метод Хафа, состоящий из следующих этапов [48]:

1. Выбор подмножества опорных точек, определяющих центры или границы связных компонент.
2. Применение преобразования Хафа для выбранного подмножества опорных точек.
3. Угол наклона, который встречается чаще всего и является предполагаемым углом наклона строк.
4. Опорные точки, формирующие между собой найденные прямые, образуют искомые строки.

Основным преимуществом метода Хафа является его точность, зависящая от частоты дискретизации. Так как анализу подвергаются только опорные точки, метод обладает хорошей скоростью работы. Однако его точность зависит от степени дискретизации ячеек накопления. Сложность метода Хафа равна $O(N^2)$.

В системе Tesseract поиск угла происходит согласно следующему алгоритму (Yср – среднее откло-

нение связных компонент, принадлежащих строке, от ее центра):

1. Применение метода связных компонент.
2. Фильтрация найденных связных компонент (цель – оставить компоненты, представляющие строки, убрать мелкие детали), остаются только компоненты высотой между 20 и 95 %.
3. Сортировка связных компонент по x-координате левой границы компоненты.
4. Первоначальное задание строк:
   4.1. $Y\text{ср} \to 0$;
   4.2. Для каждой компоненты найти текущую строку, имеющую наибольшую площадь перекрытия с текущей компонентой;
   4.3. Если такой строки нет, то
      4.3.1. Создать новую строку, прикрепить к ней компоненту, задать верхнюю и нижнюю границу строки согласно границам компоненты.
   4.4. Иначе
      4.4.1. Добавить компоненту к строке, изменить координаты строки.
   4.5. Шаг 4.1.
5. Интерполяция строк производится с помощью метода наименьших квадратов.

Данный алгоритм обладает большей точностью и скоростью работы по сравнению с методом Хафа [47]. Его сложность $O(N)$ обусловлена применением метода связных компонент и метода наименьших квадратов. В работе [51] применяется подобный подход, однако авторы дополнительно используют сформированные во время обучения модели строк для их более точного детектирования.

В работе [27] предложен алгоритм уточнения высот строк на основе анализа гистограммы высот связных компонент. Гистограмма высот найденных связных компонент сглаживается фильтром Гаусса, что значительно снижает количество пиков. В результате повторного применения фильтра Гаусса в гистограмме выделяются 3 пика, которые соответствуют высотам знаков препинания, строчных и заглавных букв соответственно. Таким образом, высоты текстовых строк могут быть уточнены на основе крайнего правого пика гистограммы.

Сегментация блока на строки основана на методе связных компонент, выявлении базовой (средней) линии и высоты строчных символов (x-height). Для символов, имеющих единообразное написание в строчном и прописном виде, эта информация является определяющей при распознавании.

5. Современные проблемы систем распознавания текстов

Наиболее сложной задачей в настоящее время является детектирование текстовой информации на изображениях и видеозаписях естественных сцен со сложным фоном в присутствии шума. В подобных случаях применение метода проекций, очевидно, не даст положительного результата. Гораздо лучше проявляют себя методы, основанные на выделении спектральных признаков, так как они являются более универсальными и устойчивыми к повороту и наличию шумов.

Немаловажной проблемой является искажение текста во время сканирования или фотографирования. При этом алгоритмы аппроксимации являются достаточно эффективными при решении подобных задач [53]. Ввиду того, что различные блоки текста могут быть подвергнуты различным типам искажений, следует проводить анализ наклона отдельно для каждого выделенного блока. Кроме того, нередко возникают разрывы и слияния символов. Страница, расположенная с нарушением границ или перекосом, создает искаженные символьные изображения, которые в дальнейшем могут быть не распознаны OCR-системой.

Реальные изображения текстов обычно далеки от совершенства, и процент ошибок распознавания для зашумленных изображений текстов часто недопустимо велик. Зашумленные изображения являются наиболее очевидной проблемой, так как даже незначительный шум может перекрывать значимые части символа или сливать соседние символы в один [52].

Таким образом, задача распознавания текстовой информации на изображениях и видеозаписях является актуальной, осложняемой необходимостью обработки информации, полученной с различных носителей. Следовательно, предложенные ранее алгоритмы работы с документами и текстовой информацией на изображениях и видео требуют существенной доработки и улучшения.

Литература

1. **Кузьмицкий, Н.Н.** Обнаружение фрагментов текста на изображениях реальных сцен на базе сверточной нейросетевой модели / Н.Н. Кузьмицкий // Информатика. – 2015. – № 2(46). – С. 12-21.
2. **Казанский, Н.Л.** Распределённая система технического зрения регистрации железнодорожных составов / Н.Л. Казанский, С.Б. Попов // Компьютерная оптика. – 2012. – Т. 36, № 3.– С. 419-428.
3. **Smith, R.W.** Hybrid page layout analysis via tab-stop detection / R.W. Smith // Proceedings of 10th International Conference on Document Analysis and Recognition (ICDAR '09). – 2009. – P. 214-245. – DOI: 10.1109/ICDAR.2009.257.
4. **Yin, X.-C.** Multi-orientation scene text detection with adaptive clustering / X.-C. Yin, W.-Y. Pei, J. Zhang, H.-W. Hao // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2015. – Vol. 37, Issue 9. – P. 1930-1937. – DOI: 10.1109/TPAMI.2014.2388210.
5. **Zuo, Z.-Y.** Multi-strategy tracking based text detection in scene videos / Z.-Y. Zuo, S. Tian, X.-C. Yin // 13th International Conference on Document Analysis and Recognition (ICDAR). – 2015. – P. 66-70. – DOI: 10.1109/ICDAR.2015.7333727.
6. **Koo, H.I.** Scene text detection via connected component clustering and nontext filtering / H.I. Koo, D.H. Kim // IEEE Transactions on Image Processing. – 2013.– Vol. 22, Issue 6. – P. 2296-2305. – DOI: 10.1109/TIP.2013.2249082.
7. **Nagy, G.** Twenty years of document image analysis in PAMI / G. Nagy // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – Vol. 22(1). – P. 38-62. – DOI: 10.1109/34.824820.
8. **Болотова, Ю.А.** Распознавание автомобильных номеров на основе метода связных компонент и иерархической временной сети / Ю.А. Болотова, В.Г. Спицын, М.Н. Ру-

домёткина // Компьютерная оптика. – 2015. – Т. 39, № 2. – С. 275-280. – DOI: 10.18287/0134-2452-2015-39-2-275-280.
9. **Jaderberg, M.** Reading text in the wild with convolutional neural networks / M. Jaderberg, K. Simonyan, A. Vedaldi, A. Zisserman // International Journal of Computer Vision. – 2016. – Vol. 116, Issue 1. – P. 1-20. – DOI: 10.1007/s11263-015-0823-z.
10. **Novikova, T.** Large-lexicon attribute-consistent text recognition in natural images / T. Novikova, O. Barinova, P. Kohli, V. Lempitsky // European Conference on Computer Vision. – 2012. – С. 752-765. – DOI: 10.1007/978-3-642-33783-3_54.
11. **Запрягаев, С.А.** Распознавание рукописных символов на основе анализа дескрипторов функций длины хорды / С.А. Запрягаев, А.И. Сорокин // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. –2009.– № 2. – С. 49-58.
12. **Глумов, Н.И.** Метод быстрой корреляции с использованием тернарных шаблонов при распознавании объектов на изображениях / Н.И. Глумов, Е.В. Мясников, В.Н. Копенков, М.А. Чичёва // Компьютерная оптика. – 2008. – Т. 32, № 3. – С. 277-282.
13. **Smith, R.W.** History of the Tesseract OCR engine: what worked and what didn't / R.W. Smith // Proceedings of SPIE. – 2013. – Vol. 8658. – 865802. – DOI: 10.1117/12.2010051.
14. **Breuel, T.M.** The OCRopus open source OCR system / T.M. Breuel // Proceedings of SPIE. – 2008. – Vol. 6815. – 68150F. – DOI: 10.1117/12.783598.
15. **Senior, A.W.** Off-line cursive handwriting recognition using recurrent neural networks / A.W. Senior // PhD thesis. – Cambridge: Cambridge University, 1994. – 121 c.
16. **Graves, A.** A novel connectionist system for unconstrained handwriting recognition / A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber // IEEE Transactions on Pattern Analysis and Machine Intelligence.– 2008. – Vol. 31, Issue 5. – P. 855-868. – DOI: 10.1109/TPAMI.2008.137.
17. **Srihari, S.N.** Document image analysis / S.N. Srihari, G.W. Zack // Proceedings of 8th International Conference on Pattern Recognition. – 1986. – P. 434-436.
18. **Гороховатский, А.В.** Детектирование текстовых областей на изображении документа методом слияния / А.В. Гороховатский // Системи обробки інформації. – 2014. – Випуск 1(117). – С. 75-81.
19. **Cattoni, R.** Geometric layout analysis techniques for document image understanding: A review [Электронный ресурс] / R. Cattoni, T. Coianiz, S. Messelodi, C.M. Modena // ITC-irst technical report TR#9703-09. – 1998. – URL: http://www.academia.edu/18416548/Geometric_Layout_Analysis_Techniques_for_Document_Image_Understanding_a_Review._TR_9703-09. – 68 p.
20. **Negi, A.** Localization, extraction and recognition of text in Telugu document images / A. Negi, K.N. Shanker, C.K. Chereddi // Proceedings of the 7-th International Conference on Document Analysis and Recognition. – 2003. – P. 1193-1197. – DOI: 10.1109/ICDAR.2003.1227846.
21. **Bukhari, S.S.** High performance layout analysis of Arabic and Urdu document images / S.S. Bukhari, F. Shafait, T.M. Breuel // Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR 2011). – 2011. – P. 1275-1279. – DOI: 10.1109/ICDAR.2011.257.
22. **Wong, K.Y.** Document analysis system / K.Y. Wong, R.G. Casey, F.M. Wahl // IBM Journal of Research and Development. – 1982. – Vol. 26(6).– P. 647-656. – DOI: 10.1147/rd.266.0647.
23. **Nagy, G.** Hierarchical representation of optically scanned documents / G. Nagy, S. Wagle // Proceedings of 7th International Conference on Pattern Recognition. – 1984.– P. 347-349.
24. **Baird, H.S.** Image segmentation by shape-directed covers / H.S. Baird, S.E. Jones, S.J. Fortune // Proceedings of 10th International Conference on Pattern Recognition. – 1990. – P. 820-825. – DOI: 10.1109/ICPR.1990.118223.
25. **Oudjemia, S.** Segmentation of complex document / S. Oudjemia, Z. Ameur, A. Ouahabi // Carpathian Journal of Electronic and Computer Engineering. – 2014. – Vol. 7(1). – P. 13-18.
26. **Breuel, T.M.** An algorithm for finding maximal whitespace rectangles at arbitrary orientations for document layout analysis / T.M. Breuel // Proceedings of the 7th International Conference on Document Analysis and Recognition. – 2003. – Vol. 1. – P. 66-70. – DOI: 10.1109/ICDAR.2003.1227629.
27. **Winder, A.** Extending page segmentation algorithms for mixed-layout document processing / A. Winder, T. Andersen, E.H.B. Smith // Proceedings of International Conference on Document Analysis and Recognition. – 2011. – P. 1245-1249. – DOI: 10.1109/ICDAR.2011.251.
28. **Breuel, T.M.** Two geometric algorithms for layout analysis / T.M. Breuel // International Workshop on Document Analysis Systems: DAS V. – 2002. – P. 188-199. – DOI: 10.1007/3-540-45869-7_23.
29. **Shafait, F.** Performance comparison of six algorithms for page segmentation / F. Shafait, D. Keysers, T.M. Breuel // International Workshop on Document Analysis Systems: DAS VII. – 2006. – P. 368-379. – DOI: 10.1007/11669487_33.
30. **Baird, H.S.** Background structure in document images / H.S. Baird // International Journal of Pattern Recognition and Artificial Intelligence. – 1994. – Vol. 8, Issue 05. – P. 1013-1030. – DOI: 10.1142/S0218001494000516.
31. **O'Gorman, L.** The document spectrum for page layout analysis / L. O'Gorman // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1993. – Vol. 15, Issue 11. – P. 1162-1173. – DOI: 10.1109/34.244677.
32. **Скворцов, А.В.** Триангуляция Делоне и её применение / А.В. Скворцов. – Томск: Изд-во Томского ун-та, 2002. – 128 с. – ISBN: 5-7511-1501-5.
33. **Kise, K.** Segmentation of page images using the area Voronoi diagram / K. Kise, A. Sato, M. Iwata // Computer Vision and Image Understanding. – 1998. – Vol. 70, Issue 3. – P. 370-382. – DOI: 10.1006/cviu.1998.0684.
34. **Mao, S.** Empirical performance evaluation methodology and its application to page segmentation algorithms / S. Mao, T. Kanungo // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2001. – Vol. 23, Issue 3. – P. 242-256. – DOI: 10.1109/34.910877.
35. **Gather, P.** Empirical performance evaluation methodology and its application to page segmentation algorithms: A review / P. Gather, A. Singh // International Journal of Advanced Research in Computer Engineering & Technology. – 2015. – Vol. 4, Issue 4. – P. 1277-1279.
36. **Esposito, F.** A knowledge-based approach to the layout analysis / F. Esposito, D. Malerba, G. Semeraro // Proceedings of the 3rd International Conference on Document Analysis and Recognition. – 1995. – Vol. 1. – P. 466-471. – DOI: 10.1109/ICDAR.1995.599037.
37. **Li, L.** Multilingual text detection with nonlinear neural network / L. Li, S. Yu, L. Zhong, X. Li // Mathematical Problems in Engineering. – 2015. – Vol. 2015. – 431608 (7 p.). – DOI: 10.1155/2015/431608.
38. **Shih, F.Y.** Adaptive document block segmentation and classification / F.Y. Shih, S.S. Chen // IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. – 1996. – Vol. 26, Issue 5. – P. 797-802. – DOI: 10.1109/3477.537322.
39. **Wang, D.** Classification of newspaper image blocks using texture analysis / D. Wang, S.N. Srihari // Computer Vision, Graphics, and Image Processing. – 1989. – Vol. 47, Issue 3. – P. 327-352. – DOI: 10.1016/0734-189X(89)90116-3.

40. **Vil'kin, A.M.** Algorithm for segmentation of documents based on texture features / A.M. Vil'kin, I.V. Safonov, M.A. Egorova // Pattern Recognition and Image Analysis. – 2013. – Vol. 23, Issue 1. – P. 153-159. – DOI: 10.1134/S1054661813010136.
41. **Sauvola, J.J.** Page segmentation and classification using fast feature extraction and connectivity analysis / J. Sauvola, M. Pietikäinen // Proceedings of the 3rd International Conference on Document Analysis and Recognition (ICDAR '95). – 1995. – Vol. 2. – P. 1127-1131. – DOI: 10.1109/ICDAR.1995.602118.
42. **Scherl, W.** Automatic separation of text, graphic and picture segments in printed material / W. Scherl, F. Wahl, H. Fuchsberger // Pattern Recognition in Practice. – 1980. – P. 213-221.
43. **Tsujimoto, S.** Major components of a complete text reading system / S. Tsujimoto, H. Asada // Proceedings of the IEEE. – 1992. – Vol. 80, Issue 7. – P. 1133-1149. – DOI: 10.1109/5.156475.
44. **Jain, A.K.** Page segmentation using texture analysis / A.K. Jain, Y. Zhong // Pattern Recognition. – 1996. – Vol. 29, Issue 5. – P. 743-770. – DOI: 10.1016/0031-3203(95)00131-X.
45. **Cattoni, R.** Geometric layout analysis techniques for document image understanding: A review [Электронный ресурс] / R. Cattoni, T. Coianiz, S. Messelodi, C.M. Modena // ITC-irst technical report TR#9703-09. – 1998. – URL: http://www.academia.edu/18416548/Geometric_Layout_Analysis_Techniques_for_Document_Image_Understanding_a_Review._TR_9703-09. – 68 p.
46. **Jain, A.K.** Text segmentation using Gabor filters for automatic document processing / A.K. Jain, S. Bhattacharjee // Machine Vision and Applications. – 1992. – Vol. 5, Issue 3. – P. 169-184. – DOI: 10.1007/BF02626996.
47. **Smith, R.** A simple and efficient skew detection algorithm via text row accumulation / R. Smith // Proceedings of the 3rd International Conference on Document Analysis and Recognition (ICDAR '95). – 1995. – Vol. 2. – P. 1145-1148. – DOI: 10.1109/ICDAR.1995.602124.
48. **U.S. Patent 3,069,654 G06K9/46, G01T5/02, G01T5/00, 382/281.** Method and means for recognizing complex patterns / P.V.C. Hough, filed of March 26, 1960, published of Desember 18, 1962.
49. **Hinds, S.C.** A document skew detection method using run-length encoding and the Hough transform / S.C. Hinds, J.L. Fisher, D.P. D'Amato // Proceedings of 10th International Conference on Pattern Recognition. – 1990. – Vol. 1. – P. 464-468. – DOI: 10.1109/ICPR.1990.118147.
50. **Rashid, S.F.** Scanning neural network for text line recognition / S.F. Rashid, F. Shafait, T.M. Breuel // 10th IAPR International Workshop on Document Analysis Systems (DAS). – 2012. – P. 105-109. – DOI: 10.1109/DAS.2012.77.
51. **Breuel, T.M.** High-performance OCR for printed English and Fraktur using LSTM networks / T.M. Breuel, A. Ul-Hasan, M.A. Al-Azawi // Proceedings of 12th International Conference on Document Analysis and Recognition. – 2013. – P. 683-687. – DOI: 10.1109/ICDAR.2013.140.
52. **Nagy, G.** Optical character recognition: An illustrated guide to the frontier / G. Nagy, T.A. Nartker, S.V. Rice // In: Proceedings of the IS&T/SPIE Symposium on Electronic Imaging: Science and Technology. – 1999. – P. 58-69.
53. **Масалович, А.** Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений [Электронный ресурс] / А. Масалович, Л. Местецкий // Труды Международной конференции «Графикон», Новосибирск. – 2006. – 4 с. – URL: http://graphicon.ru/html/2006/wr34_16_MestetskiyMasalovitch.pdf.
54. **Wang, T.** End-to-end text recognition with convolutional neural networks / T. Wang, D.J. Wu, A. Coates, A.Y. Ng, // Proceedings of 21st International Conference on Pattern Recognition (ICPR 2012). – 2012. – P. 3304-3308.
55. **Zhong, Y.** Automatic caption localization in compressed video / Y. Zhong, H. Zhang, A.K. Jain // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2000. – Vol. 22, Issue 4. – P. 385-392. – DOI: 10.1109/34.845381.

Сведения об авторах

Болотова Юлия Александровна, 1986 года рождения, в 2009 году окончила Томский политехнический университет по специальности «Программное обеспечение вычислительной техники и автоматизированных систем», доцент кафедры информационных систем и технологий. Область научных интересов: обработка изображений, распознавание образов, биологически-подобные алгоритмы. E-mail: *jbolotova@tpu.ru* .

Спицын Владимир Григорьевич, 1948 года рождения, в 1970 году окончил Томский государственный университет по специальности «Радиофизика и электроника», профессор, д.т.н., профессор Национального исследовательского Томского политехнического университета. Область научных интересов: нейронные сети, обработка изображений, распространение электромагнитных волн в случайно-неоднородных средах. E-mail: *spvg@tpu.ru* .

Осина Полина Максимовна, в 2016 окончила бакалавриат Томского политехнического университета по направлению «Информатика и вычислительная техника», обучается в магистратуре Томского политехнического университета, специализация «Компьютерный анализ и интерпретация данных». Область научных интересов: искусственные нейронные сети, фильтрация изображений и видео, разработка мобильных и web-приложений. Email: *polinaosina14@gmail.com* .

ГРНТИ: 28.23.15
Поступила в редакцию 21 февраля 2017 г. Окончательный вариант – 19 апреля 2017 г.

A REVIEW OF ALGORITHMS FOR TEXT DETECTION IN IMAGES AND VIDEOS

Yu.A. Bolotova[1], V.G. Spitsyn[1], P.M. Osina[1]
[1]Tomsk Polytechnic University, Tomsk, Russia

Abstract

This article reviews the history and state-of-the-art optical character recognition systems, such as ABBYY FineReader, Tesseract, CuneiForm, with particular attention given to their inner algorithms, including page layout analysis; page segmentation and document skew angle estimation. The overview includes the description and comparison of different methods proposed for the last 30 years in terms of speed and versatility. Critical analysis and discussions about the status of the field and open problems are reported.

Keywords: OCR, page layout analysis, text segmentation, skew detection.

Citation: Bolotova YuA, Spitsyn VG, Osina PM. A review of algorithms for text detection in images and videos. Computer Optics 2017; 41(3): 441-452. DOI: 10.18287/2412-6179-2017-41-3-441-452.

References

[1] Kuzmitskiy NN. Detection of text objects in images of real scenes based on convolutional neural network model [In Russian]. Informatics 2015; 2(46): 12-21.

[2] Kazanskiy NL, Popov SB. The distributed vision system of the registration of the railway train [In Russian]. Computer Optics 2012; 36(3): 419-428.

[3] Smith RW. Hybrid page layout analysis via tab-stop detection. Proc ICDAR'09 2009: 214-245. DOI: 10.1109/ICDAR.2009.257.

[4] Yin X-C, Pei W-Y, Zhang J, Hao H-W. Multi-orientation scene text detection with adaptive clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence 2015; 37(9): 1930-1937. DOI: 10.1109/TPAMI.2014.2388210.

[5] Zuo Z-Y, Tian S, Yin X-C. Multi-strategy tracking based text detection in scene videos. ICDAR 2015: 66-70. DOI: 10.1109/ICDAR.2015.7333727.

[6] Koo HI, Kim DH. Scene text detection via connected component clustering and nontext filtering. IEEE Trans Image Process 2013; 22(6): 2296-2305. DOI: 10.1109/TIP.2013.2249082.

[7] Nagy G. Twenty years of document image analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000; 22(1): 38-62. DOI: 10.1109/34.824820.

[8] Bolotova YuA, Spitsyn VG, Rudometkina MN. License plate recognition algorithm on the basis of a connected components method and a hierarchical temporal memory model. Computer Optics 2015; 39(2): 275-280. DOI: 10.18287/0134-2452-2015-39-2-275-280.

[9] Jaderberg M, Simonyan K, Vedaldi A, Zisserman A. Reading text in the wild with convolutional neural networks. International Journal of Computer Vision 2016; 116(1): 1-20. DOI: 10.1007/s11263-015-0823-z.

[10] Novikova T, Barinova O, Kohli P, Lempitsky V. Large-lexicon attribute-consistent text recognition in natural images. ECCV 2012: 752-765. DOI: 10.1007/978-3-642-33783-3_54.

[11] Zapryagaev SA, Sorokin AI. Handwritten character recognition based on analysis of chord-length function descriptors. Proceedings of Voronezh State University; Series: System Analysis and Information Technologies 2009; 2: 49-58.

[12] Glumov NI, Mjasnikov EV, Kopenkov VN, Chicheva MA. The method of fast correlation using ternary templates for object recognition on images [In Russian]. Computer Optics 2008; 32(3): 277-282.

[13] Smith R. History of the Tesseract OCR engine: what worked and what didn't. Proc SPIE 2013; 8658: 865802. DOI: 10.1117/12.2010051.

[14] Breuel TM. The OCRopus open source OCR system. Proc SPIE 2008; 6815: 68150F. DOI: 10.1117/12.783598.

[15] Senior AW. Off-line cursive handwriting recognition using recurrent neural networks. PhD thesis. Cambridge: Cambridge University;1994.

[16] Graves A, Liwicki M, Fernández S, Bertolami R, Bunke H, Schmidhuber J. A novel connectionist system for unconstrained handwriting recognition. IEEE Trans Pattern Anal Mach Intell 2008; 31(5): 855-868. DOI: 10.1109/TPAMI.2008.137.

[17] Srihari SN, Zack GW. Document Image analysis. Proceedings of 8-th International Conference on Pattern Recognition 1986: 434-436.

[18] Gorohovatskyi OV. The detection of text regions on image of a document using merge method. Information Processing Systems 2014; 1(117): 75-81.

[19] Cattoni R, Coianiz T, Messelodi S, Modena CM. Geometric layout analysis techniques for document image understanding: a review. ITC-irst Technical Report TR#9703-09 1998: 68p. Source: ⟨http://www.academia.edu/18416548/Geometric_p

[20] Layout_Analysis_Techniques_for_Document_Image_Understanding_a_Review._TR_9703-09⟩.

[21] Negi A, Shanker KN, Chereddi CK. Localization, Extraction and recognition of text in Telugu document images. Proc ICDAR 2003: 1193-1197. DOI: 10.1109/ICDAR.2003.1227846.

[22] Bukhari SS, Shafait F, Breuel TM. High performance layout analysis of Arabic and Urdu document images. Proc ICDAR 2011: 1275-1279. DOI: 10.1109/ICDAR.2011.257.

[23] Wong KY, Casey RG, Wahl FM. Document analysis system. IBM Journal of Research and Development 1982; 26(6): 647-656. DOI: 10.1147/rd.266.0647.

[24] Nagy G, Wagle S. Hierarchical representation of optically scanned documents. Proceedings of 7-th International conference on Pattern recognition 1984: 347-349.

[25] Baird HS, Jones SE, Fortune SJ. Image Segmentation by Shape-Directed Covers. Proc ICPR 1990: 820-825. DOI: 10.1109/ICPR.1990.118223.

[26] Oudjemia S, Ameur Z, Ouahabi A. Segmentation of complex document. Carpathian Journal of Electronic and Computer Engineering 2014; 7(1): 13-18.

[27] Breuel TM. An algorithm for finding maximal whitespace rectangles at arbitrary orientations for document layout analysis. Proc ICDAR 2003; 1: 66-70. DOI: 10.1109/ICDAR.2003.1227629.

[28] Winder A, Andersen T, Smith EHB. Extending page segmentation algorithms for mixed-layout document processing. Proc ICDAR 2011: 1245-1249. DOI: 10.1109/ICDAR.2011.251.

[29] Breuel TM. Two geometric algorithms for layout analysis. International Workshop on Document Analysis Systems 2002: 188-199. DOI: 10.1007/3-540-45869-7_23.

[30] Shafait F, Keysers D, Breuel TM. Performance comparison of six algorithms for page segmentation. International Workshop on Document Analysis Systems 2006: 368-379. DOI: 10.1007/11669487_33.

[31] Baird HS. Background structure in document images. International Journal of Pattern Recognition and Artificial Intelligence 1994; 8(05): 1013-1030. DOI: 10.1142/S0218001494000516.

[32] O'Gorman L. The document spectrum for page layout analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 1993; 15(11): 1162-1173. DOI: 10.1109/34.244677.

[33] Skvortsov AV. Delaunay trianguliation and its application [In Russian]. Tomsk: Tomsk University Publisher; 2002. ISBN: 5-7511-1501-5.
[34] Kise K, Sato A, Iwata M. Segmentation of page images using the area Voronoi diagram. Computer Vision and Image Understanding 1998; 70(3): 370-382. DOI: 10.1006/cviu.1998.0684.
[35] Mao S, Kanungo T. Empirical performance evaluation methodology and its application to page segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001; 23(3): 242-256. DOI: 10.1109/34.910877.
[36] Gather P, Singh A. Empirical performance evaluation methodology and its application to page segmentation algorithms: A review. International Journal of Advanced Research in Computer Engineering & Technology 2015; 4(4): 1277-1279.
[37] Esposito F, Malerba D, Semeraro G. A knowledge-based approach to the layout analysis. Proc ICDAR 1995; 1: 466-471. DOI: 10.1109/ICDAR.1995.599037.
[38] Li L, Yu S, Zhong L, Li X. Multilingual text detection with nonlinear neural network. Mathematical Problems in Engineering 2015; 2015: 431608. DOI: 10.1155/2015/431608.
[39] Shih FY, Chen SS. Adaptive document block segmentation and classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 1996; 26(5): 797-802. DOI: 10.1109/3477.537322.
[40] Wang D, Srihari SN. Classification of newspaper image blocks using texture analysis. Computer Vision, Graphics, and Image Processing 1989; 47(3): 327-352. DOI: 10.1016/0734-189X(89)90116-3.
[41] Vil'kin AM, Safonov IV, Egorova MA. Algorithm for segmentation of documents based on texture features. Pattern Recognition and Image Analysis 2013; 23(1): 153-159. DOI: 10.1134/S1054661813010136.
[42] Sauvola JJ, Pietikäinen M. Page segmentation and classification using fast feature extraction and connectivity analysis. Proc ICDAR '95 1995; 2: 1127-1131. DOI: 10.1109/ICDAR.1995.602118.
[43] Scherl W, Wahl F, Fuchsberger H. Automatic separation of text, graphic and picture segments in printed material. Pattern Recognition in Practice 1980: 213-221.
[44] Tsujimoto S, Asada H. Major components of a complete text reading system. Proceedings of the IEEE 1992; 80(7): 1133-1149. DOI: 10.1109/5.156475.
[45] Jain AK, Zhong Y. Page segmentation using texture analysis. Pattern Recognition 1996; 29(5): 743-770. DOI: 10.1016/0031-3203(95)00131-X.
[46] Cattoni R, Coianiz T, Messelodi S, Modena CM. Geometric layout analysis techniques for document image understanding: A review. ITC-irst Technical Report TR#9703-09 1998. Source: ⟨http://www.academia.edu/18416548/Geometric_Layout_Analysis_Techniques_for_Document_Image_Understanding_a_Review._TR_9703-09⟩.
[47] Jain AK, Bhattacharjee S. Text segmentation using Gabor filters for automatic document processing. Machine Vision and Applications 1992; 5(3): 169-184. DOI: 10.1007/BF02626996.
[48] Smith R. A simple and efficient skew detection algorithm via text row accumulation. Proc ICDAR '95 1995; 2: 1145-1148. DOI: 10.1109/ICDAR.1995.602124.
[49] Hough PVC. Method and means for recognizing complex patterns. Patent US 3069654, filed of March 26, 1960, published of Desember 18, 1962.
[50] Hinds SC, Fisher JL, D'Amato DP. A document skew detection method using run-length encoding and the Hough transform. Proc ICPR 1990; 1: 464-468. DOI: 10.1109/ICPR.1990.118147.
[51] Rashid SF, Shafait F, Breuel TM. Scanning neural network for text line recognition. 10th IAPR International Workshop on Document Analysis Systems (DAS) 2012: 105-109. DOI: 10.1109/DAS.2012.77.
[52] Breuel TM, Ul-Hasan A, Al-Azawi MA. High-performance OCR for printed English and Fraktur using LSTM networks. Proc ICDAR 2013: 683-687. DOI: 10.1109/ICDAR.2013.140.
[53] Nagy G, Nartker TA, Rice SV. Optical character recognition: an illustrated guide to the frontier. Proceedings of the IS&T/SPIE Symposium on Electronic Imaging 1999: 58-69.
[54] Masalovich A, Mestetskiy L. Warped image restoration based on continuous skeletal-border representation [In Russian]. Proceedings of the International Conference "GraphiCon" (Novosibirsk) 2006: 4 p. Source: ⟨http://graphicon.ru/html/2006/wr34_16_MestetskiyMasalovitch.pdf⟩.
[55] Wang T, Wu DJ, Coates A, Ng AY. End-to-end text recognition with convolutional neural networks. ICPR 2012: 3304-3308.
[56] Zhong Y, Zhang H, Jain AK. Automatic caption localization in compressed video. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000; 22(4): 385-392. DOI: 10.1109/34.845381.

Authors' information

Yuliya Alexandrovna Bolotova (b. 1986) graduated from Tomsk Polytechnic University in 2009, PhD associated professor at Information Systems and Technologies department, Tomsk Polytechnic University. Her research interests are image processing, object recognition, biologically-inspired models. E-mail: *jbolotova@tpu.ru* .

Vladimir Grigorievich Spitsyn (b. 1948) graduated from Tomsk State University in 1970, Radio-Physics department. He works as the Professor of Tomsk Polytechnic University. His research interests are currently focused on neural networks, image processing, electromagnetic wave propagation in random discrete media. E-mail: *spvg@tpu.ru* .

Polina Maksimovna Osina, got her bachelor's degree from Tomsk Polytechnic University in 2016 (Computer Science and Engineering), enrolled in master's of Tomsk Polytechnic University on specialization "Computer Analysis and Data Interpretation". His research interests include artificial neural networks, image and video filtration, mobile and web-applications development. E-mail: *polinaosina14@gmail.com* .

Received February 21, 2017. The final version – April 19, 2017.