

О локализации плоских объектов на изображениях со сложной структурой проективных искажений¹

Д.В. Тропин^{*,**}, Ю.А. Шемякина^{*,***}, И.А. Коноваленко^{*,****}, И.А. Фараджев^{*,***}

** ООО «Смарт Энджинс Сервис», Москва, Россия*

*** Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Россия*

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

***** Институт проблем передачи информации им. А. А. Харкевича Российской академии наук, Москва, Россия*

Поступила в редколлегию 01.06.2019

Аннотация—Статья посвящена разработке алгоритма установления собственной системы координат сложного документа на изображениях, полученных при помощи фотокамеры, на примере главного разворота паспорта РФ, две страницы которого могут не лежать в одной плоскости. Показаны особенности этой задачи, приводящие к тому, что классические методы плохо применимы. Предлагается двухэтапный подход. На первом этапе производится проективная нормализация, учитывающая ортотропную структуру страниц документа, на втором – определяются сдвиг и масштаб. Благодаря использованию разработанного алгоритма на целевом наборе данных удалось достичь сокращения числа ошибок поиска границ документа в четыре раза.

КЛЮЧЕВЫЕ СЛОВА: установление собственной системы координат документа, проективная нормализация, точки схода, детектирование отрезков, края, хребты, текстовые строки.

1. ВВЕДЕНИЕ

Автоматическое распознавание документов [1], удостоверяющих личность, т.е. извлечение персональных данных о владельце документа, является востребованной технологией, находящей свое применение в банковских системах, системах безопасности транспортных узлов, телекоммуникационных системах и т.д. Помимо собственно оцифровки текста от систем распознавания документов также требуется верное атрибутирование текстовых фрагментов по принадлежности тому или иному информационному полю (см. Рис. 1). Кроме текстовых информационных полей (имя, фамилия, отчество, пол, дата рождения и т.д.) в документах могут присутствовать и графические (подписи, печати, фотографии), которые также требуется атрибутировать. Удостоверяющие личность документы почти всегда имеют жесткую геометрическую структуру информационных полей, поэтому установление собственной системы координат на изображении для таких документов кардинально упрощает задачу атрибутирования текстовых и графических элементов.

Сложность задач распознавания изображений зависит от условий их регистрации [2]: типа устройства ввода, способа предъявления объекта съемки, параметров освещения. Так, для изображений документов, полученных со сканера, существует множество методов, успешно решающих задачу наведения, т.е. установления на изображении собственной системы координат документа [3–5]. Это объясняется, в частности, простотой геометрической модели предъявления документа: в этом случае она 3-параметрическая — неизвестен только сдвиг и поворот

¹ Работа выполнена при частичной финансовой поддержке грантов РФФИ №18-07-01387 и №18-29-26035.

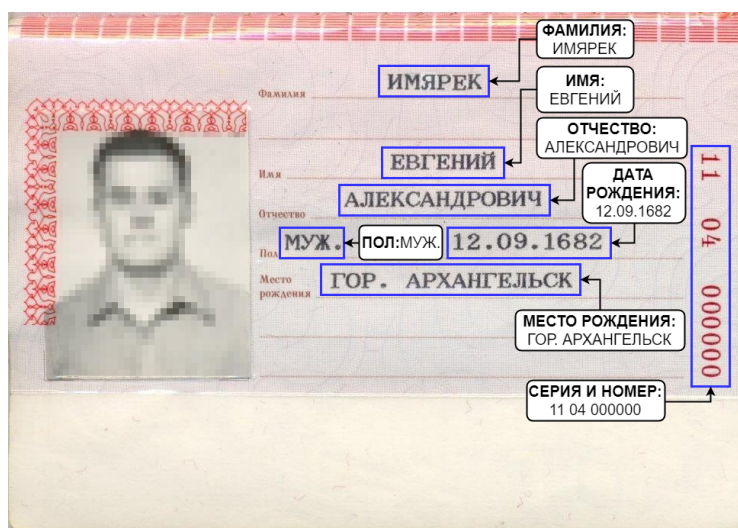


Рис. 1. Распознавание документа – оцифровка данных и атрибутирование.

образа документа. С повсеместным распространением мобильных устройств на первый план вышла задача распознавания документов по изображениям, полученным с видеокамер [1]. Если не требовать от пользователя аккуратного предъявления документа в строго определенном ракурсе, то образ документа становится подвержен проективным искажениям, что делает невозможным непосредственное применение решений, предназначенных для сканеров. Даже если не рассматривать возможные изгибы документа, при известных внутренних параметрах камеры геометрическая модель предъявления документа становится 6-параметрической. Далее будет рассматриваться именно такой случай.

В дальнейшем мы будем оценивать существующие методы наведения с точки зрения их применимости к изображениям (см. Рис. 2), содержащим главный разворот паспорта гражданина Российской Федерации (коротко, паспорта РФ). Будет показано, что существующие методы наведения для документов с жесткой геометрической структурой в таких случаях неприменимы. Поэтому в настоящей работе предлагается решать эту задачу в два этапа, а именно — проводить проективную нормализацию образа документа с точностью до неизвестных сдвига и масштаба и после этого применять уточнение с помощью 3-параметрической модели.

Аналогичные проблемы (и аналогичное их решение) возникают и для других типов документов. Например, для идентификационной карты резидента Италии, одного вида паспортов США, военного билета РФ и водительского удостоверения РФ старого образца.

В разделе 2 данной работы приведен обзор существующих методов решения задачи наведения на документ и показана их плохая применимость для решения задачи на развороте паспорта РФ. В разделе 3 рассмотрена схема предлагаемого метода наведения для многостраничных документов типа паспорта РФ и описаны ключевые алгоритмы проективной нормализации путем поиска точек схода на изображении документа. В разделе 4 приводятся экспериментальный анализ работы предложенного алгоритма.

2. МЕТОДЫ УСТАНОВЛЕНИЯ СОБСТВЕННОЙ СИСТЕМЫ КООРДИНАТ ПЛОСКОГО ОБЪЕКТА

В параграфе 2.1 описаны 3 группы существующих методов наведения. В параграфе 2.2 объясняется, почему эти методы плохо применимы к наведению на страницы разворота паспорта

кажениями и большими углами поворота успешная работа метода требует больших затрат времени.

2.2. Почему базовые методы наведения недостаточны в случае паспорта РФ

Кроме перечисленных выше недостатков методов наведения с использованием контура документа применение этих методов к наведению на разворот паспорта РФ препятствует нечеткость линии разделения страниц паспорта вследствие световых эффектов и повреждений от естественного износа (складки, разрывы и т.д.), иллюстрация которой приведена на рисунке 3.



Рис. 3. Нечеткость линии разделения 2-ой и 3-ей страниц паспорта РФ: блики на линии разделения (а); тень на линии разделения (b); складка вблизи линии разделения (c); разрыв страницы вблизи линии разделения (d).

Проанализируем теперь применимость методов, основанных на использовании особых точек. На развороте паспорта РФ есть несколько видов статических элементов (см. Рис. 4): заголовки текстовых полей, линии подчеркивания, орнамент, окаймляющий фотографию и надпись «Российская Федерация». К сожалению, заголовки полей (на второй странице это «Паспорт выдан», «Дата выдачи», «Код подразделения» и т.д., на третьей — «Фамилия», «Имя», «Отчество» и т.д.) на размытых вследствие низкого разрешения, движения и расфокусированных изображениях становятся неразличимыми, поэтому они не могут быть использованы для формирования устойчивого проективного базиса. Другой вид статических элементов — линии подчеркивания — также неприменимы для использования в методах данной группы, поскольку ни сами линии, ни их концевые точки не имеют стабильного локального окружения, позволяющего различать их между собой. Таким же недостатком обладают особые точки орнамента, окаймляющего фотографию на третьей странице. Единственной устойчивой разрешимой контрастной статической текстурой на второй странице является надпись «Российская Федерация», но, поскольку она коллиниарна, ее недостаточно для формирования устойчиво-

го проективного базиса. Таким образом, паспорт РФ содержит слишком мало разрешимых контрастных статических текстур и методы, основанные на сопоставлении созвездия особых точек, вряд ли применимы без существенных доработок.

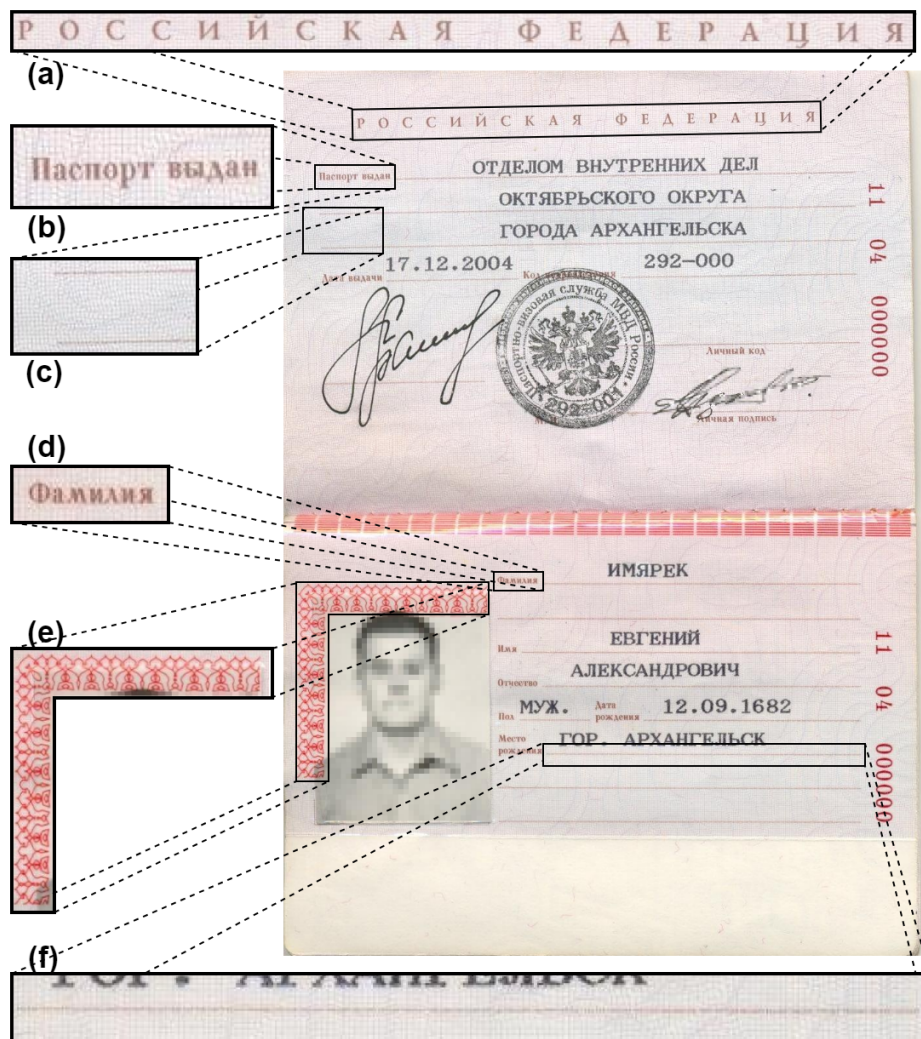


Рис. 4. Статические элементы на главном развороте паспорта РФ: надпись «Российская Федерация» (a); статические заголовки текстовых полей (b, d); крайние точки линий подчеркивания (c, f); орнамент, окаймляющий фотографию (e).

Применение методов Виолы-Джонса в случае значительных проективных искажений и произвольной ориентации документа на изображении для достижения приемлемых по качеству результатов наведения требует большого перебора вследствие классификационной природы метода.

Таким образом, ни один из базовых методов наведения не дает приемлемого по быстродействию и/или качеству результата. В [7] показано, что комбинация методов из первых двух групп может существенно улучшить результат, но для наведения на разворот паспорта РФ этого все равно недостаточно. Поэтому мы предлагаем двухэтапную схему наведения — проективная нормализация исходного изображения с последующим вычислением сдвига и масштаба методом Виолы-Джонса.

2.3. Методы проективной нормализации плоского объекта

Под задачей проективной нормализации документа здесь и далее подразумевается проективное преобразование входного изображения таким образом, чтобы документ стал подобен образцу, а окаймляющий его прямоугольник стал ортотропным.

Ключом проективной нормализации (ректификации) плоских объектов во многих существующих методах служит наличие у объектов ортотропной структуры. Здесь под ортотропной или манхэттенской структурой объекта подразумевается совокупность его протяженных физических и графических элементов, выровненных по осям его системы координат. На изображении такая структура порождает горизонтальные и вертикальные точки схода, то есть точки, в которой пересекаются осевые прямые всех горизонтальных или вертикальных элементов ортотропной структуры. Главная проблема здесь – это зашумленность прямых, точка схода которых ищется. При этом необходимо определить функционал, экстремум которого и дает оптимальную (при известных характеристиках шума) точку. Существующие подходы могут быть классифицированы по способу определения ортотропной структуры: на основе оптимизации функционала, вычисляемого по всему изображению, или по набору тех или иных отдельных протяженных элементов. Например, в широко цитируемой работе [15] предлагается метод ТИЛТ проективной нормализации, в основе которого лежит минимизация функционала (ранга матрицы), вычисляемого по изображению, в пространстве параметров проективного преобразования. Достоинством алгоритма, предложенного в этой работе, является его универсальность – метод может быть использован для решения задачи проективной нормализации объектов разной природы (см. Рис. 5). Единственным обязательным условием его применимости является наличие у прообраза регулярной ортотропной структуры. Недостаток этого алгоритма заключается в необходимости наличия априорной информации о положении зоны, в которой находится объект интереса, а также в его высокой вычислительной трудоемкости.

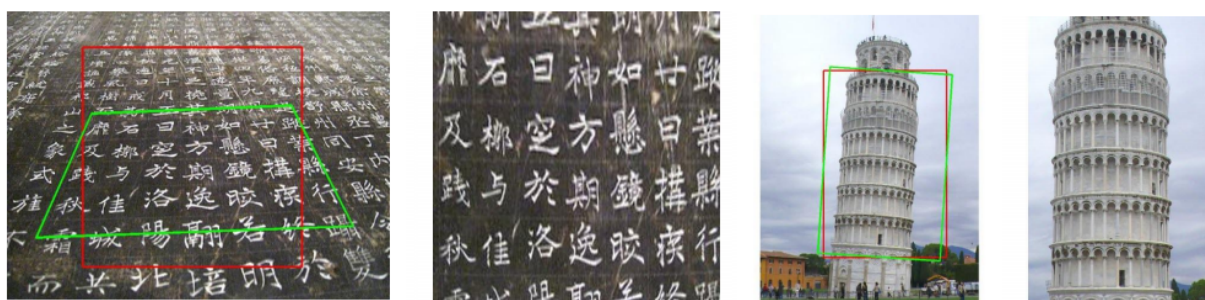


Рис. 5. Примеры работы алгоритма ТИЛТ на различных изображениях. Слева направо: китайские иероглифы, проективно нормализованное изображение китайских иероглифов, Пизанская башня, проективно нормализованное изображение Пизанской башни. Красный четырехугольник – задаваемая область интереса. Зеленый четырехугольник выровнен в соответствии с установленной системой координат объекта в области интереса.

Помимо алгоритма ТИЛТ функционалы, вычисляемые по всей области интереса, используются и в работах, посвященных задаче проективной нормализации именно текстовых объектов. Для поиска горизонтальной точки схода, соответствующей базовым линиям текстовых строк, в работах [6, 16] предложен метод, заключающийся в переборе всех возможных точек схода и выборе той, для которой проекция бинаризованного текста обладает минимальной энтропией. Недостатком данного алгоритма является высокая вычислительная сложность ввиду большого числа перебираемых вариантов. Альтернативное решение данной проблемы предложено в работах [17, 18]. В них точка схода вычисляется методом линейной регрессии, примененной

к оценкам направлений текстов в локальной окрестности каждого пикселя. Для определения направления текста в локальных окрестностях авторы используют ту же схему минимизации энтропии на проекциях.

Другой класс методов опирается на ортотропность совокупности протяженных графических элементов: границ типа край, границ типа хребет/впадина и базовых линий текстовых строк (см. Рис. 6).

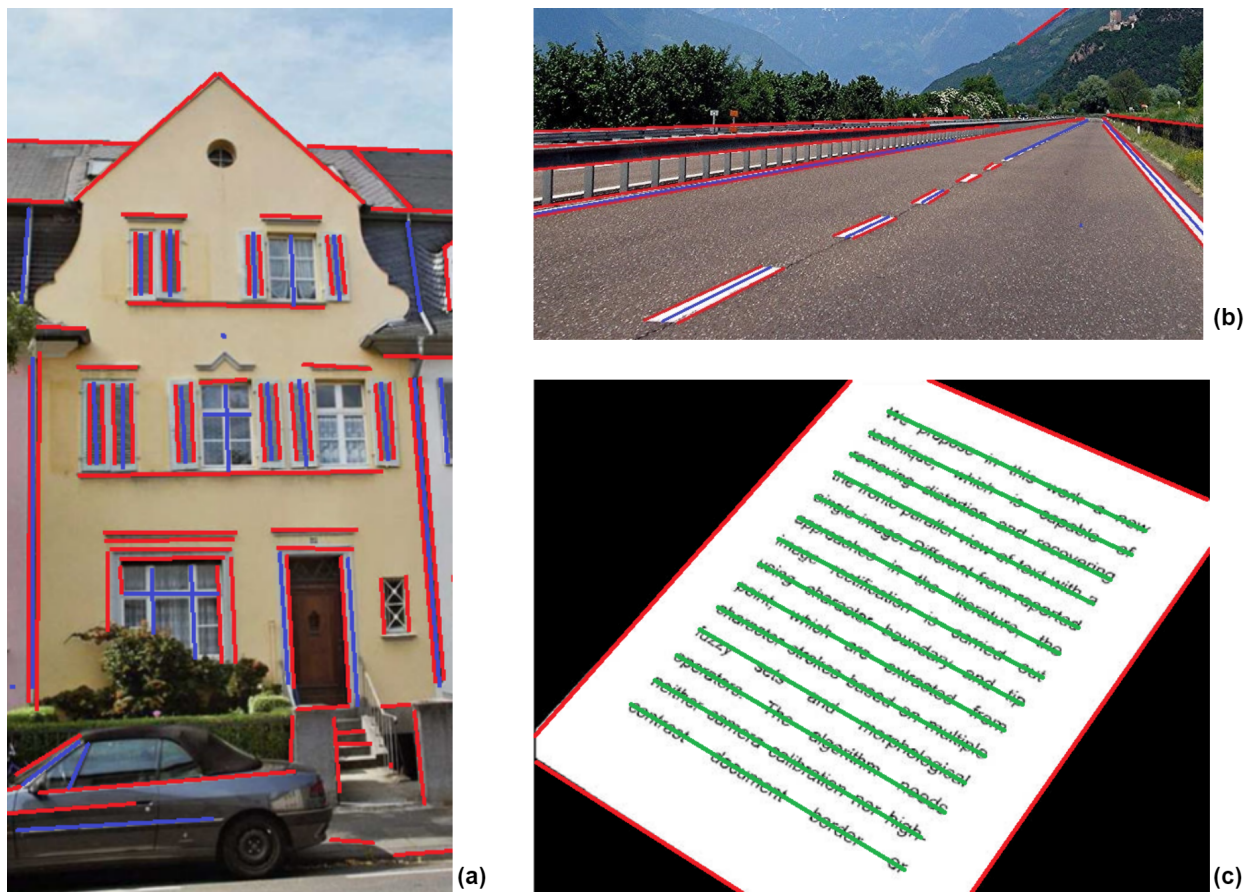


Рис. 6. Пример протяженных графических элементов на изображениях фасада дома (а), дорожного полотна (b), текстового документа (с). Отрезки, соответствующие границам типа край, выделены красным, хребты и долины – синим, центральные базовые линии текстовых строк – зеленым.

Для оценки горизонтальной точки схода на изображении текстовых объектов необходимо находить базовые линии текстовых строк. Например, в работах [19, 20] для этой цели используются центры масс компонент связности бинаризованного текста, в статье [21] ищутся локальные особенности типа пятно, а в работе [17] — краевые точки символов, на которых строятся верхняя и нижняя базовые линии строки. Для оценки положения вертикальной точки схода в части работ используются разные дополнительные предположения о структуре текста. Например, в работе [6] в зависимости от типа выравнивания текста на странице в качестве вертикальных базовых линий используется две из трех линий, соответствующих левой границе параграфов, правой границе параграфов или же центральной линии, которая существует в случае выравнивания по центру. Те же авторы в статье [16] предлагают метод вычисления точки схода уже по одной вертикальной линии в предположении о константности межстрочного интервала. В работе [21] оцениваются линии, соответствующие вертикальной

точке схода при помощи графа на локальных особенностях типа пятно и оценках принадлежности соседних «пятен» одной текстовой строке. Более низкоуровневые (локальные) характеристики используются в [19, 20]. Их авторы предлагают на первом этапе найти линии, порождаемые вертикальными образующими символов, после чего методами линейной регрессии вычислить соответствующую им точку схода. Достоинством локального подхода является отсутствие предположений о типе выравнивания текста на странице и о постоянстве межстрочного интервала. Недостатком метода является очень низкая устойчивость и неприменимость для курсива и рукописного текста.

Комбинированный подход к решению задачи ректификации текстовых документов предлагается в работах [22, 23]. В [22] авторы модифицировали оптимизируемый функционал, использующийся в алгоритме ТИЛТ, добавив в него слагаемое, в котором учитывается касание символов и базовых линий. В [23] предлагается кластеризовать полученные в результате попарного пересечения базовых линий точки схода, после чего использовать проекционные методы для выбора лучшей альтернативы.

Нейросетевой подход к решению задачи проективной нормализации изображения, содержащего текст, предложен в работах [24, 25]. Если в [24] метод рассчитан на проективную нормализацию вывесок (коротких текстовых строк), то в [25] задача решалась для изображений, содержащих образ листа формата А4. Достоинством метода, предложенного в [24], является способность к нормализации не только проективно-искаженного текста, но и изогнутого. Недостатком нейросетевых методов в контексте вычисления на мобильных процессорах является их низкое быстродействие.

Вышеперечисленные методы основываются на выделении базовых линий текстовых строк. Другой класс протяженных элементов — границы типа край, они используются для установления ортотропной структуры изображений в работах [6, 26–31]. В работе [26] предлагается алгоритм для проективной нормализации изображений плоских объектов (пластиковых карт, документов с таблицами, фасадов домов и дорожных знаков), имеющих ортотропную структуру графических и физических границ. Для установления параметров проективного преобразования авторы оптимизируют функционал, оценивающий выравнивание отрезков относительно осей изображения. Достоинством данного алгоритма является устойчивость к выбросам — отрезкам, не соответствующим ортотропной структуре объекта съемки. В работах [27, 28] рассматривается задача проективной нормализации изображения дорожной сцены, а именно, вычисление изображения, которое могло быть получено при съемке из той же точки в надир (направление, противоположное зениту). Поскольку на изображениях дорожных сцен зачастую невозможно восстановить положение обеих точек схода, используется предположение о равных ширинах дорожных полос. Благодаря этому удастся избавиться от проективных искажений [27], хотя в случае, когда калибровочная матрица камеры неизвестна, отсутствие аффинных искажений не гарантируется [28]. Более общий подход используется в работе [29], в которой ширины полос считаются произвольными, но известными. Используемый в этой работе детектор границ чувствителен как к краям, так и хребтам.

Границы типа хребет используются для нахождения ортотропных структур относительно редко. По-видимому, это связано с тем, что в большинстве работ по проективной нормализации изображений объекты съемки являются артефактами, изобилующими границами типа край. В работах, посвященных поиску дорожной разметки [32] и наведения на одномерные штрихкоды [33], встречается использование и границ типа хребет, однако в них решается не задача проективной нормализации, а задача локализации объекта.

При решении многих прикладных задач оказывается, что одного типа протяженных элементов достаточно для устойчивого формирования проективного базиса. В качестве редкого

исключения можно указать работу, в которой с целью повышения устойчивости алгоритма используется комбинация краев документа и образующих текстовых строк [6].

Однако на паспорте РФ использование только одного типа протяженных элементов не ведет к надежному решению, поскольку помимо собственно краев документа изображение паспорта РФ почти не содержит границ соответствующего типа, хребты (линии) очень тонки и потому плохо видны при низком разрешении или расфокусировке, а текстовые строки, как правило, коротки. При этом указанные потенциальные проблемы возникают независимо, а, следовательно, совместное использование всех трех типов протяженных элементов должно вести к повышению надежности алгоритма.

2.4. Методы проективной нормализации объекта, состоящего из нескольких плоскостей

Для проективной нормализации объекта, состоящего из нескольких плоскостей, например, разворота паспорта РФ, каждая из которых имеет собственную двумерную систему координат, предположения о взаимном расположении плоскостей можно классифицировать по числу их общих точек схода: когда обе точки схода общие, такая структура называется манхэттенской, когда общая точка схода одна — атлантической. Примером манхэттенской структуры является изображение плоского разворота паспорта РФ (обе страницы находятся в одной плоскости). Примером атлантической структуры (от топонима города в США, где единственная общая точка схода у границ небоскребов — вертикальная) является изображение того же разворота паспорта РФ, но уже в случае, когда вторая и третья страница находятся в разных плоскостях.

В существующих работах, объектом исследования которых являются мультипланарные объекты, предположение об атлантической структуре точек схода используется редко. Например, в статье [30] рассматривается более сложная задача восстановления трехмерной формы фасадов зданий по одному изображению, на котором у всех плоскостей есть одна общая вертикальная точка схода. В работе [31] тоже используется предположение об атлантической структуре городских сцен, но только для решения задачи калибровки камеры. Несмотря на то, что большинство снимков разворота паспорта РФ имеют атлантическую структуру, из-за складки на сгибе (см. Рис. 3.с) или разрыва (см. Рис. 3.d) две страницы могут перестать иметь общую горизонтальную точку схода. Поэтому использование предположения об атлантической структуре точек схода для паспорта РФ рискованно. Тем не менее, каждая из страниц паспорта почти всегда имеет четко выраженную манхэттенскую структуру.

3. ПРЕДЛАГАЕМАЯ СХЕМА РЕШЕНИЯ ЗАДАЧИ НАВЕДЕНИЯ

Целью данной работы является разработка вычислительно эффективного алгоритма наведения на страницы паспорта РФ на изображениях, полученных при помощи фотокамеры. Поскольку паспорт РФ, как было обсуждено выше, не удовлетворяет требованиям базовых алгоритмов установления собственной системы координат документа, предлагается использовать двухэтапную алгоритмическую схему. Она заключается в определении ортотропной структуры документа (то есть в получении двух его точек схода) и проективном преобразовании изображения к виду, когда система координат документа ортотропна в системе координат изображения, с последующим поиском страниц документа алгоритмами, используемыми при распознавании изображений со сканеров. При использовании алгоритмов наведения с собственными оценками надежности эта схема может быть дополнена перебором гипотез с различными парами точек схода.

При известных точках схода вычисление матрицы поворота и ректификация изображения не представляют существенной проблемы, эти элементы решения можно почерпнуть, например, из работ [34, 35]. Связующее звено между этими этапами (переход от матрицы поворота

к матрице проективного преобразования) тривиально [36]. Алгоритм Виолы-Джонса [14] позволяет получать оценки надежности и хорошо подходит для поиска документов на сканах [3], поэтому в данной работе он был выбран в качестве алгоритма реализации второго этапа.

Таким образом, задачу установления собственной системы координат документа можно свести к поиску пары «ортогональных» точек схода. Поскольку собственная оценка надежности алгоритма Виолы-Джонса позволяет оценить в том числе и качество ректификации, то разумно использовать нескольких кандидатов на искомую пару точек схода. Следует отметить, что применение алгоритма Виолы-Джонса, выбор наилучшей гипотезы и определение границ документа следует выполнять для каждой из n страниц документа, которые потенциально могут появиться на изображении.

Предлагаемая алгоритмическая схема изображена на рисунке 7. Здесь жирной рамкой выделены этапы проективной нормализации, к решению которых свелась задача установления собственной системы координат документа (параграфы 3.2 и 3.3). Но прежде чем перейти к описанию алгоритмов, необходимо уточнить понятие точки схода зашумленных прямых, что и будет сделано в параграфе 3.1.

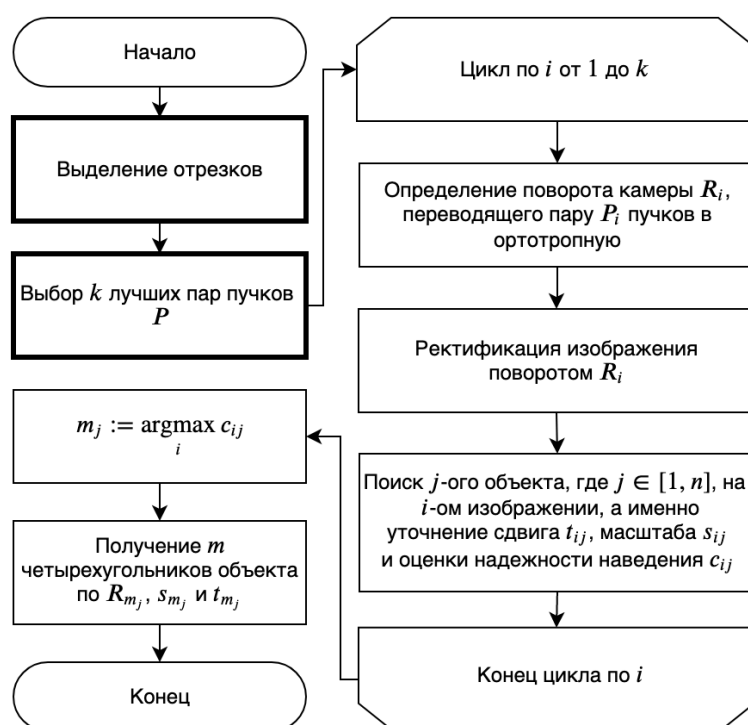


Рис. 7. Блок-схема алгоритма установления собственной системы координат документа.

3.1. Точка схода зашумленных прямых

Если прямые пересекаются в одной точке, то эта точка и называется точкой их схода. Если прямые на изображении зашумлены, то их точка схода должна минимизировать некоторый функционал, зависящий от параметров зашумления. Прямые на изображении детектируются чаще всего либо методами, связанными с преобразованием Хафа, либо проводятся через концы отрезков, идентифицированных как края, хребты или текстовые строки. Неизвестна модель зашумления прямых, полученных из преобразования Хафа. Попытка определить зашумление прямых, проведенных через две точки, зашумленные гауссовым шумом, была сделана в ра-

боте [37], однако это не привело к построению подходящего функционала. Функционалы для вычисления точек схода, используемые в работах параграфа 2.3, обладают высокой вычислительной сложностью и не имеют прозрачного физического смысла. Поэтому мы приняли решение рассматривать вместо прямых отрезки, через которые они проходят. В этом случае, используя метод максимума правдоподобия, удалось построить функционал, зависящий только от зашумления концов отрезков. Для зашумления концов используется общепринятая модель симметричного гауссова шума с нулевым средним.

Пусть $S = \{s_i = (p_1^i, p_2^i)\}$, $i = 1, \dots, N$ – коллекция отрезков, тогда оптимальная точка q (точка схода прямых, проходящих через концы отрезков) минимизирует функционал

$$f(S, q) = \sum_i \phi(s_i, q), \quad (1)$$

где $\phi(s_i, q) = \min_u (dist^2(u, p_1^i) + dist^2(u, p_2^i))$. Здесь $dist(v, p)$ – расстояние от точки p до прямой v , и минимум берется по всем прямым, проходящим через точку q . Следует отметить, что похожий подход используется в работе [26].

3.2. Выделение отрезков

Поскольку в документах рассматриваемого класса присутствуют все типы протяженных элементов (см. Рис. 8), описанных в параграфе 2.3, более того, все они дополняют друг друга, было принято решение использовать их все. Предлагаемый алгоритм выделения отрезков можно разделить на 4 части: составление карт границ 3 типов, выделение отрезков на каждой карте, объединение однотипных отрезков и объединение отрезков разных типов.

Использующийся в данной работе алгоритм выделения границ основывается на методе, предложенном Кэнни [38]: изображение сглаживается с помощью фильтра Гаусса, в каждой точке вычисляется градиент, после чего выполняется подавление немаксимумов и двухпороговая фильтрация.

Выделение границ типа край. Мы используем для границ типа край такие же условия принадлежности точки границе, что и в оригинальной статье Кэнни: точка принадлежит границе, если в ней достигается локальный максимум модуля градиента в направлении вектора градиента, а также значение модуля градиента в точке превышает заданный порог g_min , и максимальное значение модуля градиента в компоненте связности, которой принадлежит пиксель, больше, чем g_comp_min (эти пороги являются входными параметрами алгоритма).

Выделение границ типа хребет. Существующие критерии принадлежности точки границе типа хребет [39, 40] основываются на вычислении частных производных второго порядка как для фильтрации по значению, так и для установления направления хребта. В данной работе для проверки достижения локального максимума в точке предлагается не использовать условие отрицательности первого собственного значения матрицы Гессе, а сравнивать непосредственно значения яркости вблизи точки на сглаженном изображении в направлении собственного вектора, соответствующего второму собственному значению матрицы Гессе. Для повышения точности установления направления хребта при вычислении матрицы Гессе использовался дифференциальный оператор Щарра [41], обладающий вращательной симметрией. Использование исходного изображения для двухпороговой фильтрации невозможно, из-за того, что на разных частях изображения яркость фона может меняться. Для решения этой проблемы исходное изображение было стабилизировано по фону, путем деления яркости каждого пикселя исходного изображения на оценку фона cl_wing .

Выделение границ типа текст. Для получения границ типа центральные линии текстовых строк был применен тот же критерий, что и для хребтов — точка принадлежит границе,

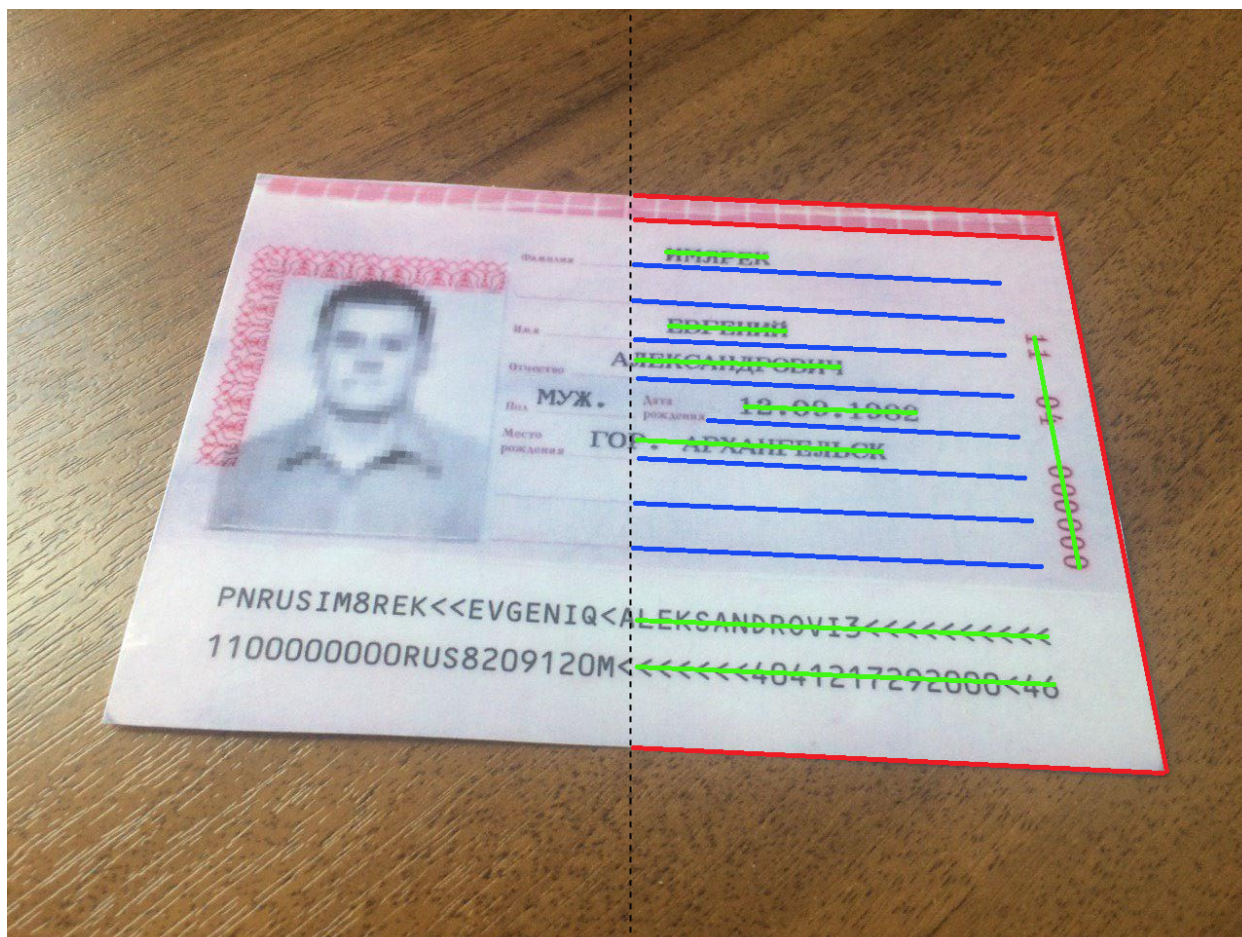


Рис. 8. Третья страница паспорта РФ. На правую половину изображения вручную нанесены отрезки трех типов: красные соответствуют отрезкам типа край, синие – хребтам, зеленые – текстам.

если в ней достигается локальный максимум яркости изображения, нормализованного по фону, в направлении второго собственного вектора матрицы Гессе. Использование того же критерия возможно, если понизить разрешение входного изображения и применить к нему фильтр Гаусса с большой дисперсией.

Выделения отрезков по карте границ. Алгоритм, использующийся для выделения отрезков по карте границ принадлежит классу локальных – отрезки собираются от точек, принадлежащих границам, последовательным присоединением к отрезкам. Как и в других алгоритмах данного класса [42, 43] его трудоемкость равна $O(W \times H)$, где W – ширина изображения, H – высота. Линейная асимптотика данной реализации достигается благодаря тому, что непосредственная сборка отрезков происходит не на «сырых» компонентах, а на так называемых треках. Компонента представляется в виде графа, вершинами которого являются точки, принадлежащие границам, и вершины соединяются ребрами, если точки являются соседними в смысле 8-связности. Трек – это граф вида простая цепь. Для установления треков был применен простой подход: сначала удаляются все вершины, в которых расстояние до ближайшего листа меньше заданного параметра алгоритма $dist_max$ и степень вершины при этом меньше или равна единице. После этого удалялись все вершины, степень которых больше двух. Тем самым из компоненты (планарного графа с циклами) получается набор треков.

Для получения отрезков по треку предлагается провести его сегментацию на участки двух типов: угловые изгибы и прямолинейные участки. Установление точек, принадлежащих из-

гибу, основано на сравнении угла между двумя прямыми, проходящими через исследуемую точку и точки того же трека, находящиеся на расстоянии $bend_dist$, с пороговым значением $angle_max$. После того как точки, принадлежащие изгибам, удалены, для оставшихся точек проверяется их принадлежность прямолинейному участку с помощью следующего критерия: вычисляются длины p и q осей эллипса, покрывающего множество точек на расстоянии str_dist от исследуемой точки, после чего отношение p/q сравнивается с порогом $rate_max$. В результате непрерывные последовательности точек трека, в которых выполняется критерий принадлежности прямолинейному участку, образуют отрезки, которые далее задаются парами концевых точек.

Объединение однотипных отрезков. Из-за того, что на первых этапах обработки границ некоторые компоненты были разбиты на несколько треков, истинные отрезки на изображении могли также распадаться на части. Для их объединения предложен подход, в основе которого лежит жадное слияние пар, полученных в процессе попарного перебора отрезков, упорядоченных по уменьшению длины. Критерий объединения заключается в выполнении следующих условий: отрезки должны быть параллельны друг другу (угол между прямыми, на которых лежат отрезки, меньше чем $angle_par_max$) и отрезки должны находиться рядом друг с другом, как в поперечном направлении (отрезки должны лежать в полосе друг друга, где ширина полосы задается параметром алгоритма $trans_max$), так и в продольном направлении (расстояние между проекциями отрезков на средней прямой должно быть меньше, чем $long_max$ от общей длины объединенного отрезка).

Последовательность описанных преобразований для границ типа край проиллюстрирована на рисунке 9.

Объединение отрезков разных типов. Целью объединения отрезков разных типов является уточнение длинных отрезков и фильтрация коротких. Использование алгоритма, описанного выше (слияние отрезков одного типа), невозможно из-за того, что отрезки разных типов редко лежат на одной прямой (они описывают физически разные границы). Для обхода этой проблемы был заменен критерий, отвечающий за то, что объединяемые отрезки находятся близко к общей прямой. Вместо проверки принадлежности отрезков полосам друг друга, вычисляются максимальные расстояния от отрезков до прямых, соответствующих противоположному отрезку, после чего это расстояние сравнивается с максимумом длин двух отрезков, умноженным на $\text{tg}(union_angle_max)$. Перебор отрезков осуществляется также попарно, от длинных к коротким. Помимо объединения проводится фильтрация коротких отрезков: устанавливается порог равный $length_min$ от средней длины всех отрезков, после чего удаляются те отрезки, длина которых меньше заданного порога.

Визуализация отрезков до и после межтипового объединения приведена на рисунке 10.

3.3. Нахождение точек схода

Входом предлагаемого алгоритма является коллекция отрезков, полученная методом, описанным в предыдущем параграфе, а результатом – пары пучков отрезков, соответствующих «ортогональным» точкам схода. В нашем случае на изображении несколько точек схода, так что не все отрезки имеют отношение к каждой из них. Поэтому мы используем робастный вариант функционала 1, где сумма берется не по всем отрезкам s , а только по тем, для которых $\phi(s = (p_1, p_2), q)$ не превышает est_max (параметр алгоритма).

Для подсчета $\phi(s = (p_1, p_2), q)$ используется метод главных компонент (частный случай ортогональной регрессии). Пусть $A = \begin{pmatrix} p_1.x - q.x & p_1.y - q.y \\ p_2.x - q.x & p_2.y - q.y \end{pmatrix}$ и $M = A^T A / 2$ – ковариационная матрица матрицы A . Тогда собственный вектор, соответствующий большему собственному

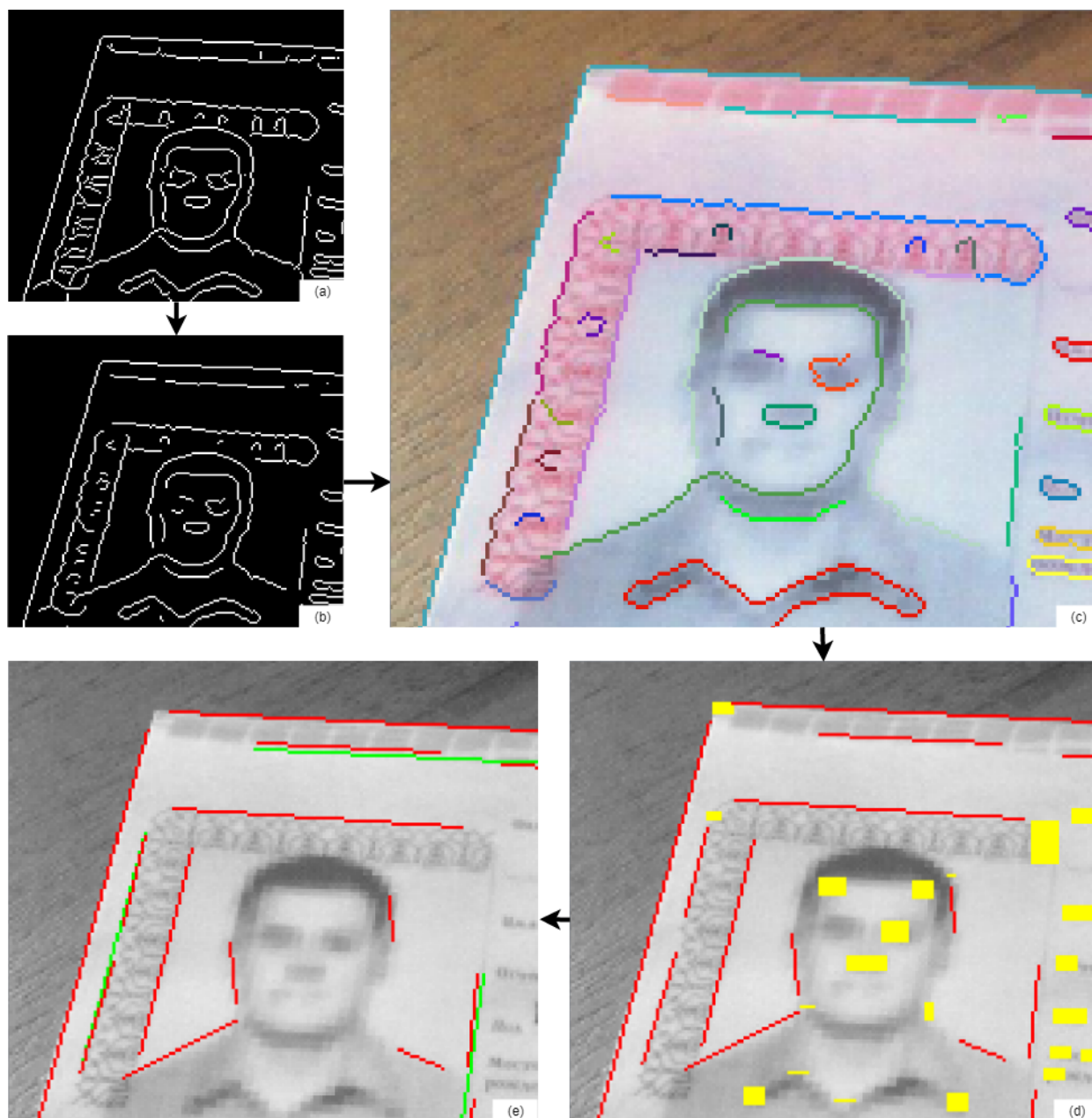


Рис. 9. Нахождение отрезков типа край на фрагменте входного изображения. Исходная карта границ (а), карта границ с отфильтрованными «хвостами» (б), треки – графы вида простая цепь (с), изгибы закрашены желтым, найденные отрезки имеют красный цвет (д), зеленым показаны объединенные отрезки (е).

значению матрицы M , определяет оптимальную прямую, а ее меньшее собственное значение равно искомой величине ϕ .

Первым делом генерируются гипотезы q точек схода – точки пересечения пары прямых, длина каждой из которых больше заданного порога $length_min$ от средней длины всех отрезков. С каждой гипотезой q связывается ее оценка $f(q)$ – значение робастного варианта функционала, пучок $P(q)$ – множество отрезков s , для которых (s, q) не превышает est_max , и $l(q)$ – суммарная длина отрезков из пучка $P(q)$. Из полученных гипотез отбираются только те,

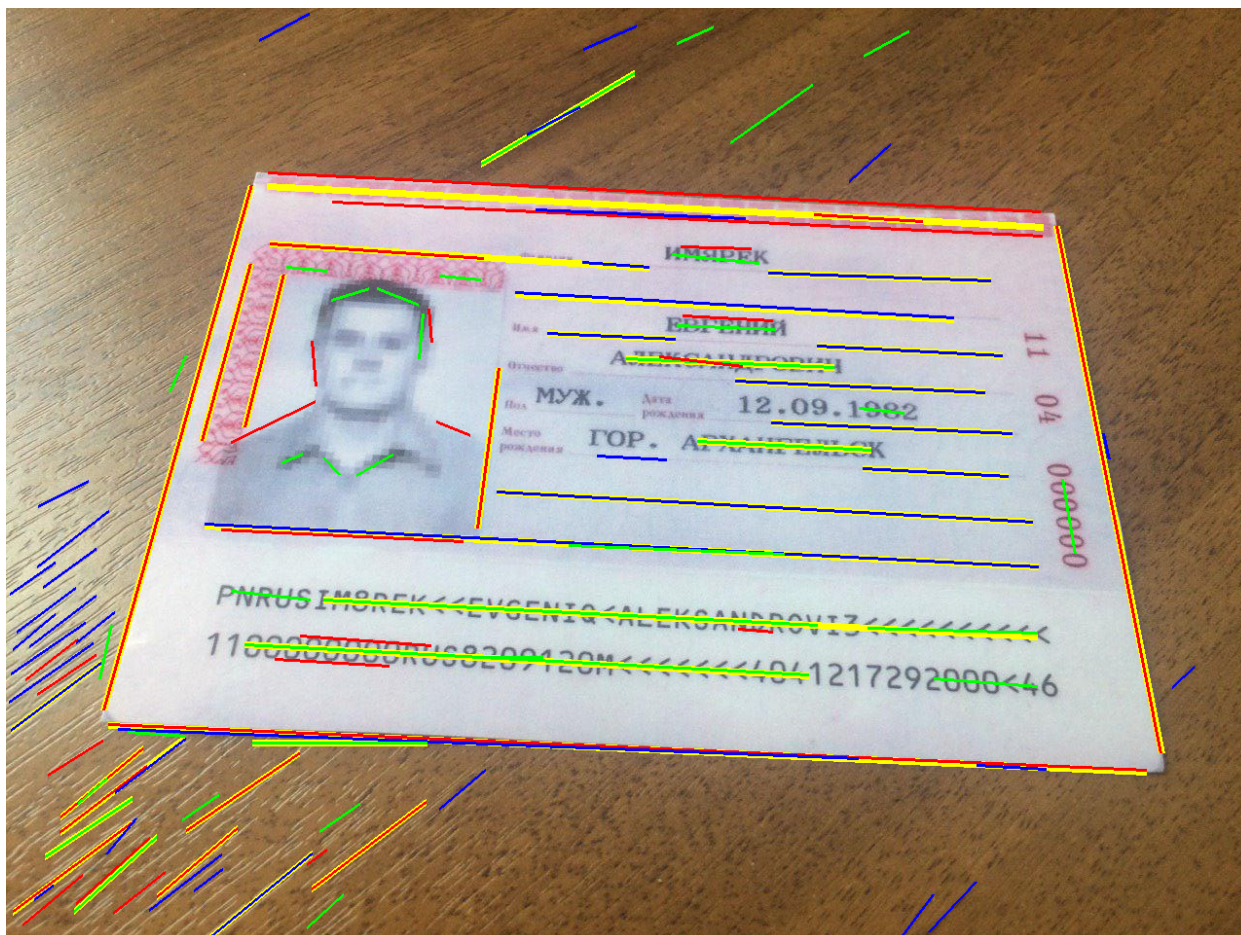


Рис. 10. Отрезки типа край, хребет и текст на стадии до объединения показаны красным, синим и зеленым соответственно. Отрезки, оставшиеся после слияния и фильтрации по средней длине, показаны желтым.

для которых минимальное расстояние до точек картинке больше заданного порога $dist_min$. Такой отбор ограничивает «степень проективности» допустимых изображений документов.

Далее гипотезы упорядочиваются по убыванию $l(q)$ и производится их кластеризация. Именно, рассматривается очередная гипотеза q и сопоставляется с каждой гипотезой q^* из уже отобранных гипотез. Если суммарная длина отрезков из теоретико-множественной разности $P(q)$ и $P(q^*)$ оказывается меньше заданного порога sim_max от средней длины всех отрезков, гипотеза q считается похожей на уже отобранную гипотезу q^* и удаляется из дальнейшего рассмотрения.

Наконец, все отобранные гипотезы уточняются методом сопряженных градиентов. После чего отбраковка гипотез по расстоянию до изображения и кластеризация повторяются.

Заключительным этапом алгоритма является отбор заданного количества k_result пар «ортогональных» особых точек. Для этого перебираются все пары отобранных точек схода и для каждой пары вычисляется угол между лучами, проходящими через центр изображения в направлении точек схода. Рассматриваемая пара считается «ортогональной», если этот угол лежит в пределах от $angle_min$ (90 градусов, меньшие углы не реализуются камерой) и $angle_max$ (ограничение на допустимую «степень проективности» изображения). В качестве результата выдается заданное количество пар пучков, соответствующих тем парам «ортогональных» точек схода, для которых суммарная длина пучков пары максимальна.

которого проиллюстрирован на графике (см. Рис. 13). Из графика видно, что оптимальное число альтернатив, подаваемых на этап уточнения сдвига и масштаба, равно 5.

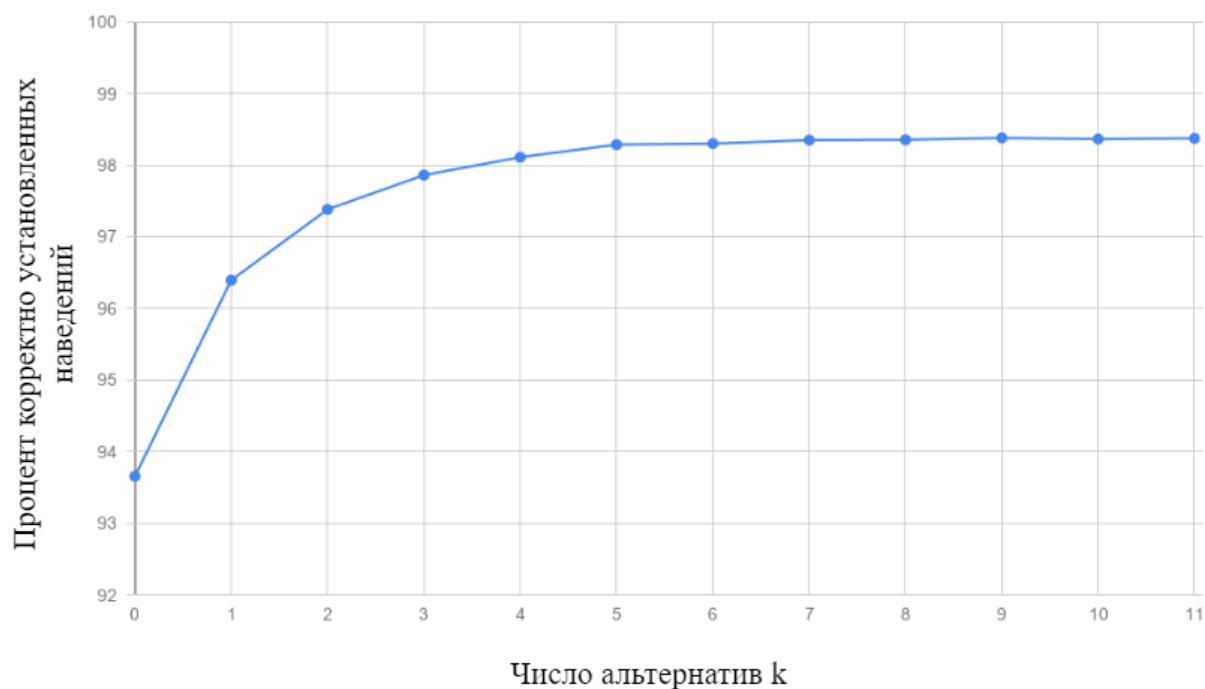


Рис. 13. График зависимости качества системы наведения (процент корректно найденных страниц) от числа подаваемых на этап установления собственной системы координат документа альтернатив. $k = 0$ соответствует случаю, когда проективная нормализация отсутствует.

Подробное сравнение качества работы системы наведения с использованием предложенного алгоритма проективной нормализации при $k = 5$ и без него приведены в таблице 1.

Таблица 1. Результаты экспериментов.

| Параметр | Без проективной нормализации | С проективной нормализацией | Изменение |
|-------------------------------------|------------------------------|-----------------------------|-----------|
| Число корректно найденных страниц | 19142 | 20089 | +947 |
| Число некорректно найденных страниц | 272 | 162 | -110 |
| Число не найденных страниц | 1025 | 188 | -837 |
| Качество | 93.65% | 98.23% | +4.63% |
| Доля ошибок | 6.35% | 1.72% | -4.63% |

Главным результатом экспериментов является то, что использование предложенного метода приводит к уменьшению числа ошибок наведения почти в 4 раза. Высокий процент хороших наведений без использования проективной нормализации объясняется тем, что сканированные документы хорошо наводятся и в этом случае.

5. ЗАКЛЮЧЕНИЕ

В настоящей работе рассмотрена задача наведения на документы с жесткой геометрической структурой в контексте распознавания документов типа паспорта РФ. Для ее решения предложена двухэтапная алгоритмическая схема, в которой от первого этапа (проективной нормализации) не требуется установления всех параметров проективного преобразования, поскольку позже они доопределяются при поиске документа на ректифицированном изображении.

В рамках решения первого этапа задачи был разработан алгоритм поиска точек схода, опирающийся на протяженные элементы трех типов: края, хребты и центральные образующие текстовых строк.

Основные результаты работы заключаются в следующем:

- Предложено искать точки схода не по линиям, а по отрезкам, через которые эти линии проходят, так что задача поиска точек схода свелась к минимизации некоторого легко вычислимого робастного функционала.
- Для проективной нормализации одновременно использовались отрезки, соответствующие трем типам границ.
- Используемые при поиске отрезков критерии были частично лишены зависимости от частных производных второго порядка. Необходимость вычисления матрицы Гессе возникла только при установлении направления хребта. Для определения локального максимума в направлении и проведения 2-пороговой фильтрации дифференцирование изображения не использовалось.
- Алгоритм проективной нормализации был внедрен в промышленную систему распознавания документов «Smart IDReader».
- На представительном тестовом наборе данных благодаря использованию предложенного алгоритма число ошибок наведения было сокращено почти в четыре раза.

Авторы благодарны В.В. Арлазарову и Д.П. Николаеву за постановку задачи, общее руководство исследованием и некоторые плодотворные идеи. Мы так же благодарны Д.П. Маталову за настройку и обучение классификаторов Виолы-Джонса.

СПИСОК ЛИТЕРАТУРЫ

1. Bulatov K., Arlazarov V.V., Chernov T., Slavin O., Nikolaev D., Smart IDReader: Document recognition in video stream. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 6, pp. 39–44.
2. Арлазаров В.В., Жуковский А.Е., Кривцов В.Е., Николаев Д.П., Полевой Д.В., Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов. *Информационные технологии и вычислительные системы*, 2014, том 3, стр. 71–81.
3. Усилин С.А., Николаев Д.П., Постников В.В., Локализация, ориентация и идентификация документов с фиксированной геометрией на изображении. *Труды Института системного анализа Российской академии наук*, 2010, том 58, стр. 248–261.
4. Усилин С.А., Николаев Д.П., Постников В.В., Быстрый алгоритм совмещения изображений документов в произвольной геометрической модели. *Труды конференции «Информационные технологии и системы»*, 2008, стр. 471–477.
5. Augereau O., Journet N., Domenger J., Semi-structured document image matching and recognition. *Document Recognition and Retrieval XX*, 2013, vol. 8658, p. 865804.
6. Clark P., Mirmehdi M., Recognising text in real scenes. *International Journal on Document Analysis and Recognition*, 2002, vol. 4, no. 4, pp. 243–257.

7. Skoryukina N., Shemiakina J., Arlazarov V., Faradjev I., Document localization algorithms based on feature points and straight lines. *Tenth International Conference on Machine Vision (ICMV 2017)*, 2018, vol. 10696, pp. 106961H.
8. Skoryukina N., Nikolaev D., Sheshkus A., Polevoy D., Real time rectangular document detection on mobile devices. *Seventh International Conference on Machine Vision (ICMV 2014)*, 2015, vol. 9445, pp. 94452A.
9. Zhukovsky A., Nikolaev D., Arlazarov V., Postnikov V., Polevoy D., Skoryukina N., Chernov T., Shemiakina J., Mukovozov A., Konovalenko I., Segments Graph-Based Approach for Document Capture in a Smartphone Video Stream. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1, pp. 337–342.
10. Javed K., Shafait F., Real-time Document Localization in Natural Images by Recursive Application of a CNN. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1, pp. 105–110.
11. Puybareau É., Géraud T., Real-Time Document Detection in Smartphone Videos. *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 1498–1502.
12. Awal A.M., Ghanmi N., Sicre R., Furon T., Complex Document Classification and Localization Application on Identity Document Images. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1, pp. 426–431.
13. Fischler M.A., Bolles R.C., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981, vol. 24, no. 6, pp. 381–395.
14. Viola P., Jones M., Rapid object detection using a boosted cascade of simple features. *CVPR (1)*, 2001, vol. 1, pp. 511–518.
15. Zhang Z., Ganesh A., Liang X., Ma Y., TILT: Transform Invariant Low-Rank Textures. *International Journal of Computer Vision*, 2012, vol. 99, no. 1, pp. 1–24.
16. Clark P., Mirmehdi M., Rectifying perspective views of text in 3D scenes using vanishing points. *Pattern Recognition*, 2003, vol. 36, no. 11, pp. 2673–2686.
17. Lu S., Chen B.M., Ko C.C., Perspective rectification of document images using fuzzy set and morphological operations. *Image and Vision Computing*, 2005, vol. 23, no. 5, pp. 541–553.
18. Liang J., DeMenthon D., Doermann D., Geometric rectification of camera-captured document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, vol. 30, no. 4, pp. 591–605.
19. Merino-Gracia C., Mirmehdi M., Sigut J., González-Mora J.L., Fast perspective recovery of text in natural scenes. *Image and Vision Computing*, 2013, vol. 31, no. 10, pp. 714–724.
20. Wang Y., Sun Y., Liu C., Layout and Perspective Distortion Independent Recognition of Captured Chinese Document Image. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1, pp. 591–596.
21. Pilu M., Extraction of illusory linear clues in perspectively skewed documents. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, vol. 1.
22. Kil T., Koo H.I., Cho N.I., Scene text rectification using glyph and character alignment properties. *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3663–3668.
23. Simon C., In K., Correcting geometric and photometric distortion of document images on a smartphone. *Journal of Electronic Imaging*, 2015, vol. 24, no. 1, p. 013038.
24. Shi B., Wang X., Lyu P., Yao C., Bai X., Robust scene text recognition with automatic rectification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4168–4176.
25. Abbas S.A., Recovering homography from camera captured documents using convolutional neural networks. *arXiv preprint arXiv:1709.03524*, 2017.

26. An J., Koo H.I., Cho N.I., Rectification of planar targets using line segments. *Machine Vision and Applications*, 2017, vol. 28, no. 1-2, pp. 91–100.
27. Nieto M., Salgado L., Plane rectification through robust vanishing point tracking using the Expectation-Maximization algorithm. *2010 IEEE International Conference on Image Processing*, 2010, pp. 1901–1904.
28. He Q., Chu C.H., Lane Detection and Tracking through Affine Rectification. *MVA*, 2007, pp. 536–539.
29. Krokhhina D., Blinov V., Gladilin S., Tarhanov I., Postnikov V., Fast roadway detection using car cabin video camera. *Eighth International Conference on Machine Vision (ICMV 2015)*, 2015, vol. 9875, pp. 98751F.
30. Zaheer A., Rashid M., Khan S., Shape from angle regularity. *European Conference on Computer Vision*, 2012, pp. 1–14.
31. Lee H., Shechtman E., Wang J., Lee S., Automatic upright adjustment of photographs with robust camera calibration. *IEEE transactions on pattern analysis and machine intelligence*, 2013, vol. 36, no. 5, pp. 833–844.
32. López A., Serrat J., Canero C., Lumbreras F., Graf T., Robust lane markings detection and road geometry computation. *International Journal of Automotive Technology*, 2010, vol. 11, no. 3, pp. 395–407.
33. Ando S., Hontani H., Automatic visual searching and reading of barcodes in 3-D scene. *IVEC2001. Proceedings of the IEEE International Vehicle Electronics Conference 2001. IVEC 2001 (Cat. No. 01EX522)*, 2001, pp. 49–54.
34. Caprile B., Torre V., Using vanishing points for camera calibration. *International journal of computer vision*, 1990, vol. 4, no. 2, pp. 127–139.
35. Шемякина Ю.А., Жуковский А.Е., Коноваленко И.А., Николаев Д.П., Алгоритм автоматического кадрирования цифровых изображений при проективном преобразовании. *Труды Института системного анализа Российской академии наук*, 2014, том. 68, № S1, стр. 142–149.
36. Zhang Z., Flexible camera calibration by viewing a plane from unknown orientations. *Iccv*, 1999, vol. 99, pp. 666–673.
37. Шемякина Ю.А., Использование точек и прямых для вычисления проективного преобразования по двум изображениям плоского объекта. *Информационные технологии и вычислительные системы*, 2017, № 3, стр. 79–91.
38. Canny J., A computational approach to edge detection. *Readings in computer vision*, 1987, pp. 184–203.
39. Lindeberg T., *Scale-space theory in computer vision*, 2007, pp. 2495–2504.
40. Steger C., An unbiased detector of curvilinear structures. *IEEE Transactions on pattern analysis and machine intelligence*, 1998, vol. 20, no. 2, pp. 113–125.
41. Jähne B., Scharr H., Körkel S., *Principles of filter design. Handbook of Computer Vision and Applications*. London: Academic Press, 1999.
42. Von Gioi R.G., Jakubowicz J., Morel J., Randall G., LSD: a line segment detector. *Image Processing On Line*, 2012, vol. 2, pp. 35–55.
43. Akinlar C., Topal C., EDLines: A real-time line segment detector with a false detection control. *Pattern Recognition Letters*, 2011, vol. 32, no. 13, pp. 1633–1642.

Localization of planar objects on the images with complex structure of projective distortion

Tropin D.V., Shemiakina J.A., Konovalenko I.A., Faradjev I.A.

Abstract—In this article an algorithm is constructed for camera-based complex document localization, using as an example the main two-page spread of Russian national passport with possibly noncoplanar pages. Properties of this task are described, and it is shown that the classical methods are insufficient. A two-stage approach is proposed. The first stage performs projective normalization taking into account an orthotropic structure of document pages. The second stage determines the shift and scale components of the required transformation. Experimental results show that using the developed algorithm the number of document localization errors is reduced by four times on a reference dataset.

KEYWORDS: localization, projective normalization, image rectification, vanishing points, segment detection, edge, ridge, text texture.