

Среднеквадратичная невязка координат как критерий точности нормализации изображений при оптическом распознавании документов¹

И.А. Коноваленко

*Институт проблем передачи информации, Российская академия наук, Москва, Россия
ООО Смарт Энджинс Сервис, Москва, Россия*

Поступила в редколлегию 05.08.2020

Аннотация—Оптическое распознавание текста на сделанном с произвольного ракурса изображении требует предварительной нормализации изображения – преобразования к такому виду, как если бы оно было получено с удобного для распознавания ракурса. При этом неточности нормализации приводят к ошибкам распознавания. На сегодняшний день в литературе предложен ряд критериев точности нормализации, однако их соответствие качеству распознавания не исследуется. В данной работе для случая документа фиксированной структуры введена нормальная вероятностная модель распознавания, в соответствии с которой вероятность верного распознавания символа падает по гауссиане с ростом невязки координат этого символа. Для этой модели доказано, что критерий точности нормализации изображения, равный среднеквадратичной по текстовым полям документа невязке координат, монотонно связан с вероятностью верного распознавания всего документа. Наконец, среднеквадратичная невязка координат для важнейшего случая проективной нормализации и состоящей из прямоугольников области интереса впервые были вычислена аналитически.

КЛЮЧЕВЫЕ СЛОВА: нормализация изображений, проективное преобразование, оптическое распознавание символов, распознавание документов, критерии точности, невязка координат.

ВВЕДЕНИЕ

Нормализация изображений

Изображения одного и того же объекта сцены существенно отличаются при различных условиях его съёмки: способе съёмки, ракурсе, освещении, взаимном расположении частей объекта, прочих объектах сцены, оптических свойствах среды и т.д. Это естественное явление значительно усложняет анализ изображений. Поэтому в случае, когда условия съёмки контролируемы, они как правило выбираются *нормальными*, т.е. априори удобными для последующего анализа формируемого изображения. Обыденный пример: использование сканера для формирования изображений документов. В случае же, когда условия съёмки неконтролируемы, остаётся возможность контролировать их виртуально. А именно, изображение подвергается *нормализации* [1, 2, 3] – преобразованию к виду, которое оно могло бы иметь, будучи сделанным при нормальных условиях съёмки. В случае плоского объекта важнейшим пунктом нормальных условий съёмки является нормальность оптической оси камеры к плоскости объекта [4, 5].

¹ Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научных проектов №18-29-26035 и №17-29-03514.

Полученные при произвольных условиях съёмки изображения часто рассматриваются как результат *искажения* воображаемых изображений, полученных при нормальных условиях съёмки. Нормализация в таком случае понимается как устранение (коррекция, компенсация) этих искажений [5, 6, 7, 8].

Если нормальные условия съёмки единственны, то нормализация является идемпотентной операцией: двукратное применение нормализации к изображению имеет тот же результат, что и однократное. Для случая, когда нормальные условия съёмки не единственны [9, 10], разумно потребовать идемпотентность дополнительно. Тогда нормализация не изменит изображение, изначально сделанное при нормальных условиях съёмки. Именно идемпотентность является общим свойством нормализации изображений и нормализации векторов.

Нормализацию изображений можно классифицировать по типу применяемых к изображению преобразований на цветовую [11], геометрическую [2, 3, 12] и общую. Цветовая нормализация подразумевает, что значение пикселя меняется независимо от его координат. Её не следует путать с растяжением гистограмм изображений, которое также называется нормализацией [13]. Геометрическая нормализация сводится к преобразованию координат пикселей. Общая нормализация не подразумевает ограничений на преобразования изображений.

Нормализация изображений всегда опирается на некоторую модель целевого объекта сцены. Например, такой моделью может быть *эталонное изображение* [9, 14] – изображение аналогичного объекта, полученное при нормальных условиях съёмки. В таком случае геометрическая нормализация изображения может быть реализована как его совмещение (image registration [15]) с эталонным изображением.

Далее в этой работе будем рассматривать только геометрическую нормализацию, называя её для краткости также нормализацией. Для её реализации используются следующие классы преобразований координат: изометрия [9, 16, 17], аффинные [18], полиномиальные [19], дробно-полиномиальные [20], центрально-проективные [4, 5], проективные [21, 22, 23, 24, 25] и произвольные [26, 27].

Нормализация применяется в качестве предварительной обработки изображений в следующих задачах их анализа: поиск похожих изображений в базах данных [28, 29, 30], распознавание телевизионных передач [31], детекция неплоских участков сцены [32], анализ медицинских изображений [6], дистанционное зондирование Земли (ДЗЗ) [5, 19, 33] (в этом случае нормализацию также принято называть ортокоррекцией, ортотрансформированием, орторектификацией, ортоисправлением), распознавание автомобильных номеров [9], детекция лиц [26] и их распознавание [34], классификация типов документов [35] и их распознавание [4, 36, 37, 38]. Кроме того, нормализация изображения документа применима для облегчения восприятия его человеком [39].

Обзор критериев точности геометрической нормализации изображений

Рассмотрим возможные интерфейсы алгоритмов нормализации. На вход подаётся подлежащее нормализации изображение I_{input} . Помимо него, в той или иной форме на вход могут подаваться априорная информация о целевом объекте, и информации о том, какие условия съёмки считаются нормальными. Кроме того, некоторые алгоритмы нормализации используют независимые данные об условиях съёмки, например об освещении и ракурсе [5, 7, 32, 40]. Возвращает алгоритм нормализации преобразование координат пикселей \hat{H} (или его параметры), которое необходимо применить к входному изображению I_{input} для получения алгоритмически нормализованного изображения I_{alg} (см. пример на рис. 1). Непосредственное преобразование изображения либо не производится вовсе (когда достаточно знать преобразование), либо ввиду своей нетривиальности отдаётся на реализацию специально предназначенным для этого алгоритмам. Дополнительно может возвращаться оценка некоторых условий съёмки.

Все алгоритмы нормализации работают не точно. Для того чтобы формализовать точность алгоритмически нормализующего преобразования \hat{H} , экспертно задаётся идеально нормализующее преобразование H (ground truth). Тогда преобразование \hat{H} может рассматриваться как оценка преобразования H . Изображение, полученное в результате применения \hat{H} к входному изображению I_{input} , назовём идеально нормализованным изображением I_{ideal} . Естественно, что нормализация тем точнее, чем ближе преобразование \hat{H} к преобразованию H или чем ближе изображение I_{alg} к изображению I_{ideal} . Однако определить такую близость можно бесконечным количеством неравносильных способов, т.е. выбор критерия точности нормализации влияет на то, какой алгоритм нормализации будет признан наиболее точным. В литературе предложено большое количество критериев точности нормализации. Прежде чем рассмотреть их, введём необходимые дополнительные обозначения.

Через $\mathbf{r} \stackrel{\text{def}}{=} [x \ y]^T$ обозначим декартовы координаты пикселей на плоскости изображения I_{ideal} и определим остаточное искажение

$$V \stackrel{\text{def}}{=} \hat{H}H^{-1}, \quad (1)$$

для каждой точки сцены переводящее координаты \mathbf{r} её образа на изображении I_{ideal} в координаты $V(\mathbf{r})$ её образа на изображении I_{alg} . В случае точной работы алгоритма нормализации V является тождественным преобразованием. Введём также невязку координат [41] (см. пример на рис. 2):

$$d(\mathbf{r}) \stackrel{\text{def}}{=} \|\mathbf{r} - V(\mathbf{r})\|_2, \quad (2)$$

показывающую расстояние, на которое образ точки сцены на изображении I_{alg} оказался смещён относительно образа той же точки сцены на изображении I_{ideal} .

В ряде случаев самими нормальными условиями съёмки определено, какая область изображения I_{ideal} представляет интерес (например, содержит образ целевого объекта). Тогда назовём её областью интереса и обозначим $R \subset \mathbb{R}^2$. В противном случае примем за область интереса R всё изображение I_{ideal} . Образ области интереса R на плоскости изображения I_{alg} обозначим $Q \stackrel{\text{def}}{=} V[R] \stackrel{\text{def}}{=} \{V(\mathbf{r}) : \mathbf{r} \in R\}$.

Теперь перейдём непосредственно к обзору критериев точности геометрической нормализации изображений. В некоторых работах предлагается оценивать точность нормализации визуально [12, 20, 27]. Формальные же критерии точности можно разделить на три типа: косвенные, яркостные и геометрические. Ниже рассмотрим все три типа отдельно.

Косвенные критерии

Выше было показано, что нормализация применяется как этап предобработки изображений при решении разнообразных задач анализа изображений. Соответственно, существует подход, при котором точность нормализации прямо определяется как качество решения задачи, в котором она применяется. Например, в работах [23, 36, 37, 42, 43] в качестве точности нормализации изображения документа выбрано качество распознавания текста на алгоритмически нормализованном изображении I_{alg} , в [35] – точность классификации документов, а в [31] – доля верно распознанных телевизионных передач. Так определённая точность зависит от нормализации лишь косвенно, поэтому соответствующие критерии точности мы будем называть косвенными. Этот подход противоречит принципу модульности ПО, который требует, чтобы разработка (а значит и тестирование) алгоритмов нормализации не зависело от прочих алгоритмов [27].

Яркостные критерии

Яркостными критериями точности нормализации мы будем называть критерии, при вычислении которых учитываются яркости изображений I_{alg} и I_{ideal} . Например, в статьях [44, 45, 46,

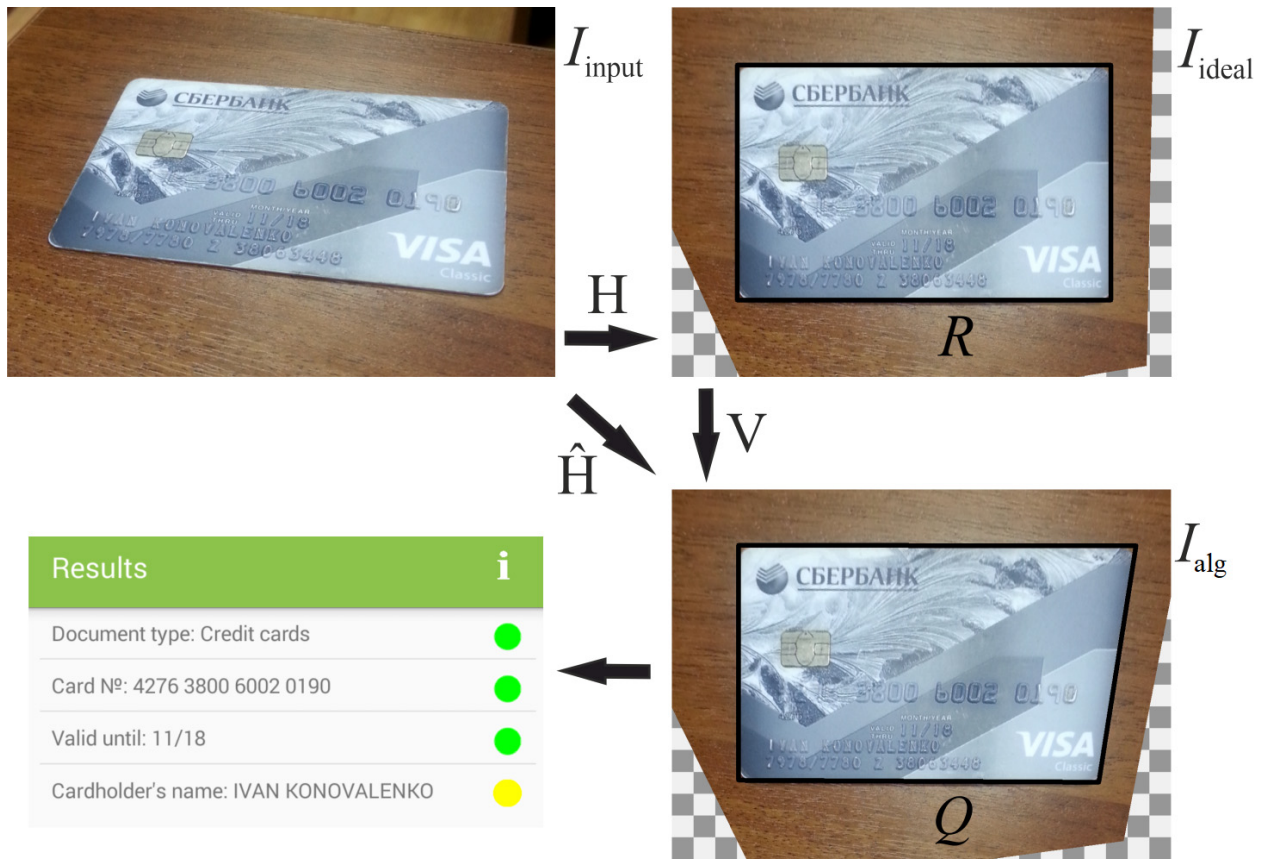


Рис. 1. Общая схема преобразований, где I_{input} – полученное с произвольного ракурса изображение документа, I_{ideal} – идеально нормализованное изображение, I_{alg} – алгоритмически нормализованное изображение и его распознавание.

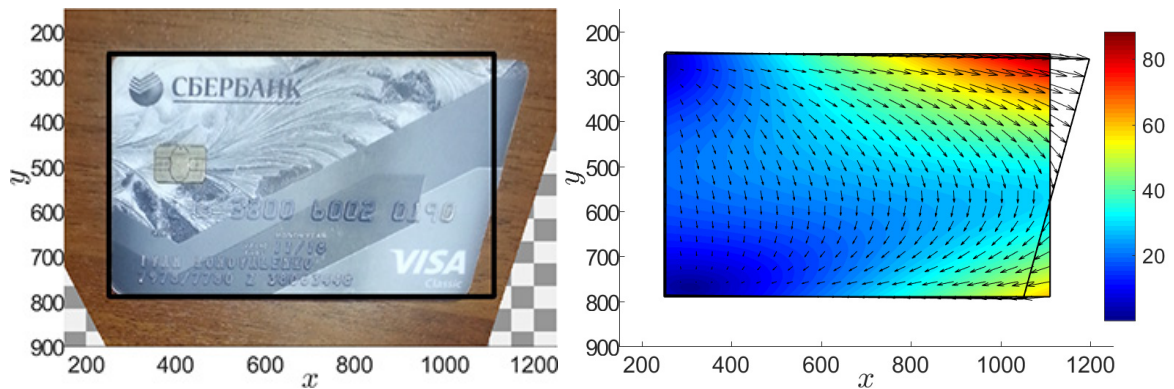


Рис. 2. Невязка координат. Слева: алгоритмически нормализованное изображение I_{alg} , рамкой ограничена область интереса R . Справа: векторное поле смещений остаточного искажения $V(\mathbf{r}) - \mathbf{r}$, $\mathbf{r} \in R$, цветом показана невязка координат $d(\mathbf{r})$.

47] в качестве точности нормализации используется среднеквадратичная поточечная разность яркостей этих изображений:

$$I_2(I_{alg}, I_{ideal}; R) \stackrel{\text{def}}{=} \sqrt{\frac{1}{S(R)} \int_R \|I_{alg}(\mathbf{r}) - I_{ideal}(\mathbf{r})\|_2^2 d\mathbf{r}}, \quad (3)$$

где $S(R)$ – площадь области интереса R , в работе [8] – пиковое отношение сигнал-шум, рассчитанное между ними, в статье [9] – мера соответствия распределений их яркостей, а в статьях [48, 49, 50] – метрика Васерштейна.

Недостаток яркостных критериев заключается в том, что на их значения влияет яркость входного изображения I_{input} : при одинаковой нормализации различающихся только яркостью входных изображений значения яркостного критерия также будут различаться.

Геометрические критерии

Геометрическими критериями точности нормализации мы будем называть критерии, зависящие только от идеальной нормализующего преобразования H , его алгоритмической оценки \hat{H} и области интереса R , а также от производных величин: Q , V , d . Геометрические критерии точности являются естественными для геометрической нормализации и чаще других встречаются в литературе. Перечислим их:

1. Близость параметров, задающих преобразования \hat{H} и H [18, 46]. Проблемой так определённой точности представляется отсутствие каких-либо её интерпретаций.
2. Коэффициент Жаккара [51], равный площади пересечения множеств Q и R , отнесенной к площади их объединения:

$$K_{\text{Jaccard}}(Q, R) \stackrel{\text{def}}{=} \frac{S(Q \cap R)}{S(Q \cup R)}. \quad (4)$$

Он использовался, например, на конкурсе «Smartphone document capture» конференции ICDAR [52]. В работе [53] предложена его более подходящая для оптимизации модификация.

3. Метрика Хаусдорфа, равная наибольшему расстоянию от точек одного множества до соответствующих им ближайших точек второго множества:

$$d_{\text{Hausdorff}}(Q, R) \stackrel{\text{def}}{=} \max\left\{\max_{\mathbf{q} \in Q} \min_{\mathbf{r} \in R} \|\mathbf{q} - \mathbf{r}\|_2, \max_{\mathbf{r} \in R} \min_{\mathbf{q} \in Q} \|\mathbf{r} - \mathbf{q}\|_2\right\}. \quad (5)$$

Она использовалась для сопоставления объектов [54], для выравнивания частично заслоненных контуров [29], для робастной детекции лиц [26], для вычисления близости двух изображений [16]. В работах [28, 33] опубликованы её модификации.

4. Расстояние Фреше, которое определяется следующим образом:

$$F(Q, R) \stackrel{\text{def}}{=} \inf_{a, b} \max_{t \in [0, 1]} \|\delta Q(a(t)), \delta R(b(t))\|_2, \quad (6)$$

где δQ и δR – непрерывные отображения отрезка $[0, 1]$ в границы множеств Q и R соответственно, a, b – непрерывные неубывающие сюръекции отрезка $[0, 1]$ в себя (репараметризация). От метрики Хаусдорфа расстояние Фреше отличается тем, что накладывает ограничения на повторное использование точек из Q и R . Этот критерий использовался для измерения близости двух контуров после проективного выравнивания в задаче проективно-инвариантного распознавания плоских замкнутых контуров [55].

5. Среднеквадратичная невязка координат

$$L_2(d; R) \stackrel{\text{def}}{=} \sqrt{\frac{1}{S(R)} \int_R d^2(\mathbf{r}) d\mathbf{r}} \quad (7)$$

использовалась как критерий точности нормализации в задачах устранения радиальной дисторсии [56], создания панорам [45, 57, 58], совмещения космических снимков [19, 59], анализа медицинских изображений [6], распознавания текста [60, 61]. Известно также использование средней (не среднеквадратичной) невязки координат [24].

6. Максимальная невязка координат (минимаксный критерий):

$$L_{\infty}(d; R) \stackrel{\text{def}}{=} \max_{\mathbf{r} \in R} d(\mathbf{r}) \quad (8)$$

использовалась как критерий точности нормализации в задачах привязки космических снимков [59], детекции лиц [26], распознавания текста [24, 61, 62, 63].

7. Максимальная невязка направлений была предложена в работе [64] для формализации точности нормализации изображений текстовых документов.
8. Некоторые критерии предназначены исключительно для случая, когда преобразования \hat{H} и H – проективные, а область интереса R – прямоугольник. Тогда её образ Q – четырёхугольник. Например, в работах [4, 65] для формализации точности нормализации используется ошибка в соотношении сторон четырёхугольника Q , в работе [7] – угол между противоположными сторонами четырёхугольника Q и их правильная ориентированность, в статье [5] – отношение минимального и максимального углов четырёхугольника Q , в работе [43] – суммарная близость длин противоположных сторон Q .

В действительности вышеперечисленные критерии не всегда применялись в точности так, как это было описано. В некоторых случаях вместо идеально нормализованного изображения использовалось эталонное изображение, а цель состояла в поиске оптимальной нормализации по соответствующему критерию, а не в тестировании найденной нормализации.

Интересно отметить, что все перечисленные критерии точности зависят от \hat{H} и H только через $V = \hat{H}H^{-1}$ (1). Поэтому точность оценки \hat{H} преобразования H можно понимать как близость остаточного искажения V к тождественному преобразованию.

Таким образом, в литературе предложено большое количество критериев точности нормализации. Однако вопрос их ориентированности на решение тех или иных задач анализа изображений не исследуется, хотя для разных задач с точки зрения качества их решения предпочтительнее могут оказаться разные алгоритмы нормализации. Это ставит под сомнение соответствие разрабатываемых алгоритмов нормализации стоящим перед ними задачам. Настоящая работа направлена на заполнение этого пробела.

1. ОБОСНОВАНИЕ СРЕДНЕКВАДРАТИЧНОЙ НЕВЯЗКИ КООРДИНАТ КАК КРИТЕРИЯ ТОЧНОСТИ ГЕОМЕТРИЧЕСКОЙ НОРМАЛИЗАЦИИ

В этом разделе мы рассмотрим задачу распознавания документов фиксированной структуры и покажем, что при некоторых простых модельных предположениях о распознавании среднеквадратичная невязка координат (7) является оптимальным критерием точности нормализации изображений с точки зрения качества последующего распознавания.

1.1. Нормальная вероятностная модель распознавания документов фиксированной структуры

Будем говорить, что документ имеет *фиксированную структуру*, если его текстовое содержание сгруппировано в текстовые поля с заранее известным положением на документе. К документам фиксированной структуры можно отнести банковские карты (см. рис. 1), водительские удостоверения, национальные и заграничные паспорта, страховые свидетельства, свидетельства о рождении, идентификационные карты, пластиковые пропуска и т.д. Априорное знание структуры документа позволяет значительно повысить качество его распознавания.

Алгоритм оптического распознавания документов фиксированной структуры получает на вход алгоритмически нормализованное изображение I_{alg} документа и возвращает распознанное текстовое содержание его полей (см. рис. 1). Точность нормализации при этом существенно

влияет на качество распознавания. Далее сформулируем модель, описывающую вероятность верного распознавания каждого символа. Произвольному символу документа поставим в соответствие его координаты \mathbf{r} на изображении I_{ideal} . На изображении I_{alg} им соответствуют координаты $V(\mathbf{r})$. Так как структура документа фиксирована и известна, для алгоритма распознавания координаты \mathbf{r} известны и имеют смысл ожидаемого положения символа, а координаты $V(\mathbf{r})$ – неизвестного фактического положения символа. Значит невязка координат $d(\mathbf{r}) = \|\mathbf{r} - V(\mathbf{r})\|_2$ (2) для символа с координатами \mathbf{r} соответствует расстоянию между его ожидаемым и действительным положениями на нормализованном изображении I_{alg} . Тогда следует ожидать, что вероятность p верного распознавания символа уменьшается с ростом невязки d этого символа. Более конкретно примем, что символ, имеющий невязку d , распознаётся верно с вероятностью:

$$p(d) = p_0 \exp\left(-\frac{1}{2} \frac{d^2}{\sigma^2}\right), \quad (9)$$

где вероятность верного распознавания в отсутствии невязки $p_0 = p(0)$ и масштаб невязки σ – неизвестные параметры (рис. 3), причём распознавание символов происходит независимо. Такую вероятностную модель распознавания будем называть *нормальной*.

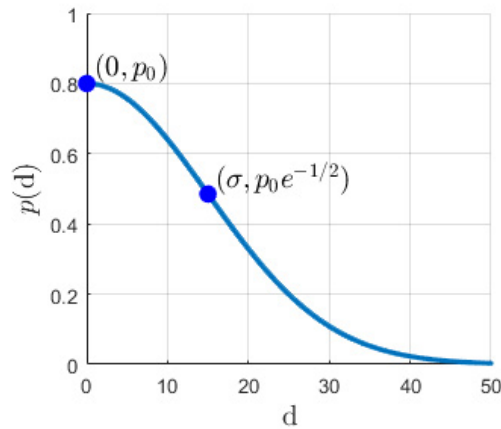


Рис. 3. Модельная зависимость вероятности p верного распознавания символа от невязки координат d для этого символа.

1.2. Постановка задачи построения критерия точности нормализации

Будем считать, что качество алгоритма распознавания (включая нормализацию) тем выше, чем больше вероятность P верного распознавания всего документа. Тогда для тестирования нормализации подойдёт такой критерий точности L , что будет связан с вероятностью P монотонно убывающе. Выполнение такого требования влечёт увеличение качества распознавания с уменьшением значения L .

1.3. Построение критерия точности нормализации

Так как распознавание символов происходит независимо, вероятность верного распознавания всего документа $P = \prod_{i=1}^n p_i$, где $p_i = p(d_i)$ – вероятности верного распознавания отдельных

символов. Используя нормальную вероятностную модель алгоритма распознавания (9), получаем:

$$P = \prod_{i=1}^n p_0 \exp\left(-\frac{1}{2} \frac{d_i^2}{\sigma^2}\right) = p_0^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n d_i^2\right). \quad (10)$$

Взяв в качестве критерия L среднеквадратичную невязку по всем символам документа $\sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2}$, имеем монотонно убывающую зависимость P от L : $P = p_0^n \exp\left(-\frac{n}{2\sigma^2} L^2\right)$, что решает поставленную задачу при любых значениях параметров p_0 и σ . Так как текстовые поля включают в себя те и только те области изображения, где могут находиться символы, то вместо $\sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2}$ будем использовать $\sqrt{\frac{1}{S(R)} \int_R d^2(\mathbf{r}) d\mathbf{r}}$, где область интереса R в данном случае – множество точек изображения I_{ideal} , на котором находятся текстовые поля. Таким образом, искомым критерием точности нормализации является упоминаемая в обзоре среднеквадратичная невязка координат (7):

$$L = L_2(d; R) \stackrel{\text{def}}{=} \sqrt{\frac{1}{S(R)} \int_R d^2(\mathbf{r}) d\mathbf{r}}.$$

2. ВЫЧИСЛЕНИЕ СРЕДНЕКВАДРАТИЧНОЙ НЕВЯЗКИ КООРДИНАТ ПРОЕКТИВНОГО ПРЕОБРАЗОВАНИЯ

Во многих случаях целевой объект можно считать плоским, а камеру – камерой-обскурой [66]. Тогда полученные с произвольных ракурсов изображения целевого объекта оказываются связаны проективными преобразованиями [67, 68], вследствие чего обоснованной геометрической нормализацией является нормализация проективная, т.е. такая, при которой преобразования координат пикселей \tilde{H} и H , а значит и V (1) – проективны. Проективная нормализация является наиболее часто используемой геометрической нормализацией изображений.

Обоснованный в предыдущем разделе критерий точности нормализации (7) для случая проективной нормализации с учётом формулы (2) математически представляет из себя среднеквадратичную невязку координат проективного преобразования (СНКПП) V на множестве R :

$$L_2(V; R) = \sqrt{\frac{1}{S(R)} \int_R \|\mathbf{r} - V(\mathbf{r})\|_2^2 d\mathbf{r}}. \quad (11)$$

Ранее СНКПП уже использовалась для тестирования алгоритмов нормализации, например в работах [45, 58, 60], но вычислялась численно. В этом разделе мы впервые вычислим её аналитически для важнейшего случая, когда область интереса R состоит из прямоугольников.

2.1. Границы применимости

Проективное преобразование V параметризуем однородной матрицей (матрицей гомографии) $V \stackrel{\text{def}}{=} (v_{ij}) \in \mathbb{R}^{3 \times 3}$ следующим образом:

$$V(\mathbf{r}) \stackrel{\text{def}}{=} \frac{\begin{bmatrix} v_{11}x + v_{12}y + v_{13} \\ v_{21}x + v_{22}y + v_{23} \\ v_{31}x + v_{32}y + v_{33} \end{bmatrix}}{v_{31}x + v_{32}y + v_{33}}. \quad (12)$$

Введём функцию

$$Z(\mathbf{r}) \stackrel{\text{def}}{=} v_{31}x + v_{32}y + v_{33}. \quad (13)$$

Прямую на плоскости изображения I_{ideal} с уравнением $Z(\mathbf{r}) = 0$ будем называть горизонтом. Допустим, что область интереса R не лежит строго по одну сторону от горизонта. Точки на горизонте обращают знаменатель преобразования (12) в ноль, что соответствует их бесконечной удалённости на плоскости входного изображения I_{input} , и не могут поэтому присутствовать на кадре в силу его конечных размеров. Физически такие точки сцены находятся под углом обзора $\pi/2$ камеры. Точки же по разные стороны от горизонта не могут одновременно присутствовать на изображении I_{input} , т.к. точки по одну из сторон имеют для камеры угол обзора $> \pi/2$, т.е. находятся позади неё. Таким образом, по крайней мере часть области интереса отсутствует на входном изображении I_{input} . В таком случае формула (11) теряет смысл критерия точности. Поэтому мы будем рассматривать только те случаи, когда область интереса лежит строго по одну сторону от горизонта:

$$\begin{cases} Z(\mathbf{r} \in R) < 0, \\ Z(\mathbf{r} \in R) > 0. \end{cases} \tag{14}$$

Это условие также обеспечивает математическую корректность формулы СНКПП (11).

2.2. Случай ортотропной прямоугольной области интереса

Для приложений анализа изображений важен случай, когда область интереса R является ортотропным прямоугольником:

$$R = [x_1, x_2] \times [y_1, y_2], \quad x_1 < x_2, \quad y_1 < y_2.$$

Тогда площадь области интереса $S(R) = (x_2 - x_1)(y_2 - y_1)$ и

$$L_2(\mathbf{V}; R) = \sqrt{\frac{1}{S(R)} \int_{x_1}^{x_2} \int_{y_1}^{y_2} \left\| \begin{bmatrix} x - \frac{v_{11}x + v_{12}y + v_{13}}{v_{31}x + v_{32}y + v_{33}} \\ y - \frac{v_{21}x + v_{22}y + v_{23}}{v_{31}x + v_{32}y + v_{33}} \end{bmatrix} \right\|_2^2 dx dy}. \tag{15}$$

Введём первообразную

$$F(x, y) \stackrel{\text{def}}{=} \int \int \left\| \begin{bmatrix} x - \frac{v_{11}x + v_{12}y + v_{13}}{v_{31}x + v_{32}y + v_{33}} \\ y - \frac{v_{21}x + v_{22}y + v_{23}}{v_{31}x + v_{32}y + v_{33}} \end{bmatrix} \right\|_2^2 dx dy, \tag{16}$$

тогда по теореме Ньютона – Лейбница:

$$L_2(\mathbf{V}; R) = \sqrt{\frac{1}{S(R)} F(x, y)|_{x_1}^{x_2}|_{y_1}^{y_2}}, \tag{17}$$

где

$$F(x, y)|_{x_1}^{x_2}|_{y_1}^{y_2} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \tag{18}$$

Теперь для вычисления (11) остаётся найти первообразную F :

$$F(x, y) = F_1(x, y) + F_2(x, y), \tag{19}$$

$$\begin{aligned}
F_1(x, y) &\stackrel{\text{def}}{=} \iint \left(x - \frac{v_{11}x + v_{12}y + v_{13}}{v_{31}x + v_{32}y + v_{33}} \right)^2 dx dy = \frac{1}{18(v_{31}v_{32})^3} \times \\
&\times \left[9 \left(2b_{12}z_1(x) - b_{15}y - 4b_{13}v_{32} \right) b_{15}y - 18 \left(b_{12}z_1(x) - b_{13}v_{32} \right)^2 l(\mathbf{r}) \right. \\
&+ 6 \left(v_{32}^2(6b_{11}b_{13} + b_{14}y + 2b_{15}y^2) - v_{32}z_1(x)(b_{14} + 2b_{15}y) + 2b_{15}z_1(x)^2 \right) Z(\mathbf{r})l(\mathbf{r}) \\
&\left. - \left(v_{32}^2(9b_{11}(4b_{13} + b_{12}y) + v_{32}(9b_{13} + 4b_{12}y)y) - 6v_{32}z_1(x)(b_{14} + b_{15}y) + 12b_{15}z_1(x)^2 \right) v_{32}y \right] \\
&+ \left(\frac{1}{3}x^2 + \left(\frac{v_{11}}{v_{31}} \right)^2 \right) xy - (v_{11}v_{31}x + b_{12}y + 2b_{13}) \frac{xy}{v_{31}^2},
\end{aligned} \tag{20}$$

$$\begin{aligned}
F_2(x, y) &\stackrel{\text{def}}{=} \iint \left(y - \frac{v_{21}x + v_{22}y + v_{23}}{v_{31}x + v_{32}y + v_{33}} \right)^2 dx dy = \frac{1}{18(v_{32}v_{31})^3} \times \\
&\times \left[9 \left(2b_{22}z_2(y) - b_{25}x - 4b_{23}v_{31} \right) b_{25}x - 18 \left(b_{22}z_2(y) - b_{23}v_{31} \right)^2 l(\mathbf{r}) \right. \\
&+ 6 \left(v_{31}^2(6b_{21}b_{23} + b_{24}x + 2b_{25}x^2) - v_{31}z_2(y)(b_{24} + 2b_{25}x) + 2b_{25}z_2(y)^2 \right) Z(\mathbf{r})l(\mathbf{r}) \\
&\left. - \left(v_{31}^2(9b_{21}(4b_{23} + b_{22}x) + v_{31}(9b_{23} + 4b_{22}x)x) - 6v_{31}z_2(y)(b_{24} + b_{25}x) + 12b_{25}z_2(y)^2 \right) v_{31}x \right] \\
&+ \left(\frac{1}{3}y^2 + \left(\frac{v_{22}}{v_{32}} \right)^2 \right) yx - (v_{22}v_{32}y + b_{22}x + 2b_{23}) \frac{yx}{v_{32}^2},
\end{aligned} \tag{21}$$

где

$$\begin{aligned}
b_{11} &\stackrel{\text{def}}{=} v_{11} + v_{33}, & b_{21} &\stackrel{\text{def}}{=} v_{22} + v_{33}, \\
b_{12} &\stackrel{\text{def}}{=} v_{31}v_{12} - v_{11}v_{32}, & b_{22} &\stackrel{\text{def}}{=} v_{32}v_{21} - v_{22}v_{31}, \\
b_{13} &\stackrel{\text{def}}{=} v_{31}v_{13} - v_{11}v_{33}, & b_{23} &\stackrel{\text{def}}{=} v_{32}v_{23} - v_{22}v_{33}, \\
b_{14} &\stackrel{\text{def}}{=} 3(b_{11}b_{12} + b_{13}v_{32}), & b_{24} &\stackrel{\text{def}}{=} 3(b_{21}b_{22} + b_{23}v_{31}), \\
b_{15} &\stackrel{\text{def}}{=} v_{32}b_{12}, & b_{25} &\stackrel{\text{def}}{=} v_{31}b_{22}, \\
z_1(x) &\stackrel{\text{def}}{=} v_{31}x + v_{33}, & z_2(y) &\stackrel{\text{def}}{=} v_{32}y + v_{33},
\end{aligned} \tag{22}$$

$$l(\mathbf{r}) \stackrel{\text{def}}{=} \log Z(\mathbf{r}). \tag{23}$$

Хотя СНКПП и является вещественной, в процессе её вычисления могут возникать комплексные числа. Это обстоятельство представляет из себя серьёзное неудобство, особенно при программной реализации. Комплексные числа возникают в случае отрицательности функции Z , которая встречается под логарифмом (23). Но ввиду ограничения (14) функция Z может принимать либо только положительные, либо только отрицательные значения. Из определения Z (13) ясно, что сменить её знак на положительный можно умножив матрицу V на -1 . Такая замена корректна, т.к. матрица V однородна: определяемое ей по формуле (12) проективное преобразование не меняется от её умножения на любое ненулевое число. Поэтому для избежания комплексных чисел достаточно выполнить $V := -V$, если $Z(\mathbf{r}) < 0$ для одной произвольно выбранной точки области интереса $\mathbf{r} \in R$.

Таким образом, аналитически вычислена СНКПП на ортотропной прямоугольной области интереса.

2.3. Случай прямоугольной области интереса

Теперь обобщим полученное выше решение на случай, когда R – произвольно ориентированный прямоугольник. Представим его как образ поворота U ортотропного прямоугольника R_0 :

$$R = U(R_0), \quad R_0 = [x_1, x_2] \times [y_1, y_2], \quad x_1 < x_2, \quad y_1 < y_2,$$

тогда $S(R) = S(R_0)$. Введём новые координаты $\mathbf{p} \stackrel{\text{def}}{=} U^{-1}(\mathbf{r})$, тогда $\mathbf{r} = U(\mathbf{p})$ и

$$\begin{aligned}
 L_2(V; R) &\stackrel{\text{def}}{=} \sqrt{\frac{1}{S(R)} \int_R \|\mathbf{r} - V(\mathbf{r})\|_2^2 d\mathbf{r}} = \\
 &= \sqrt{\frac{1}{S(R_0)} \int_{R_0} \|U(\mathbf{p}) - V \circ U(\mathbf{p})\|_2^2 d\mathbf{p}} = \\
 &= \sqrt{\frac{1}{S(R_0)} \int_{R_0} \|\mathbf{p} - U^{-1} \circ V \circ U(\mathbf{p})\|_2^2 d\mathbf{p}} = \\
 &= L_2(U^{-1} \circ V \circ U; R_0).
 \end{aligned} \tag{24}$$

Таким образом, СНКПП V на прямоугольной области интереса $R = U(R_0)$, где U – поворот, R_0 – ортогольный прямоугольник, равняется СНКПП $U^{-1} \circ V \circ U$ на R_0 , вычисление которой было описано в предыдущем подразделе.

2.4. *Случай состоящей из прямоугольников области интереса*

Теперь рассмотрим случай, когда область интереса состоит из прямоугольников:

$$R = \cup_{i=1}^n R_i: \quad R_i \cap R_j = \emptyset, \quad i \neq j.$$

Тогда $S(R) = \sum_{i=1}^n S(R_i)$ и

$$\begin{aligned}
 L_2(V; R) &\stackrel{\text{def}}{=} \sqrt{\frac{1}{S(R)} \int_R \|\mathbf{r} - V(\mathbf{r})\|_2^2 d\mathbf{r}} = \\
 &= \sqrt{\frac{1}{S(R)} \sum_{i=1}^n \int_{R_i} d^2(\mathbf{r}) d\mathbf{r}} = \\
 &= \sqrt{\frac{1}{S(R)} \sum_{i=1}^n \frac{S(R_i)}{S(R_i)} \int_{R_i} d^2(\mathbf{r}) d\mathbf{r}} = \\
 &= \sqrt{\sum_{i=1}^n \frac{S(R_i)}{S(R)} L_2^2(V; R_i)}.
 \end{aligned} \tag{25}$$

Таким образом, аналитическое вычисление СНКПП на состоящей из прямоугольников области интереса свелось к полученному в предыдущем подразделе аналитическому вычислению СНКПП на прямоугольной области интереса.

3. ЗАКЛЮЧЕНИЕ

В работе введена математическая модель системы распознавания документов фиксированной структуры по изображению, полученному с произвольного ракурса. В рамках введённой модели доказано, что критерий точности нормализации изображения, равный среднеквадратичной по текстовым полям документа невязке координат, монотонно связан с вероятностью последующего верного распознавания всего документа. Наконец, среднеквадратичная невязка координат для важнейшего случая проективной нормализации и состоящей из прямоугольников области интереса впервые были вычислена аналитически.

Полученные результаты могут быть использованы как для построения алгоритмов нормализации изображений, так и для дальнейшего развития критериев точности нормализации изображений.

СПИСОК ЛИТЕРАТУРЫ

1. Путятин Е. П., Прокопенко Д. О., Печеная Е. М. Вопросы нормализации изображений при проективных преобразованиях. *Радиоэлектроника и информатика*, 1998, Т. 2, № 3, стр. 82–86.
2. Любченко Валентин Анатольевич, Путятин Евгений Петрович. Математические модели разложения проективных преобразований в задачах нормализации. *Радиоэлектроника и информатика*, 2002, № 2 (19), стр. 57–59.
3. Чехлов Д. О., Абламейко С. В. Нормализация изображений относительно перспективного преобразования на основе геометрических параметров. *Информатика*, 2004, № 3, стр. 67–76.
4. Rodríguez-Piñeiro J., Comesaña-Alfaro P., Pérez-González F., Malvido-García A. A new method for perspective correction of document images. *Document Recognition and Retrieval XVIII*. International Society for Optics and Photonics, 2011. vol. 787410, pp. 1–12.
5. Холопов И. С. Алгоритм коррекции проективных искажений при маловысотной съёмке. *Компьютерная оптика*, 2017, Т. 41, № 2, стр. 284–290.
6. Baltzopoulos V. A videofluoroscopy method for optical distortion correction and measurement of knee-joint kinematics. *Clinical Biomechanics*, 1995, vol. 10, no. 2, pp. 85–92.
7. Calore E., Pedersini F., Frosio I. Accelerometer based horizon and keystone perspective correction. *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*. IEEE, 2012, pp. 205–209.
8. Цветков В. Ю. Геометрические модели многоракурсных изображений и проективная компенсация движения камеры. *Доклады Белорусского государственного университета информатики и радиоэлектроники*, 2014, Т. 86, № 8, стр. 41–47.
9. Мурыги К. В. Нормализация изображения автомобильного номера и сегментация символов для последующего распознавания. *Штучный интеллект*, 2010, № 3, стр. 364–369.
10. Николаев П. П. Проективно инвариантное описание овалов с симметриями трех родов. *Вестник РФФИ*, 2016, № 4 (92), стр. 38–54, DOI: 10.22204/2410-4639-2016-092-04-38-54.
11. Gong H., Finlayson G. D., Fisher R. B. and Fang F. 3D color homography model for photo-realistic color transfer re-coding. *The Visual Computer*, 2019, vol. 35, no. 3, pp. 323–333, DOI: 10.1007/s00371-017-1462-x.
12. Clark A. J., Green R. D., Grant R. N. Perspective correction for improved visual registration using natural features. *Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference*, 2008, pp. 1–6.
13. Gonzalez Rafael C, Woods Richard E. *Digital Image Processing*. Prentice Hall, 2008.
14. Ваничев А. Ю. Нормализация силуэтов объектов в системах технического зрения. *Программные продукты и системы*, 2007, № 3, стр. 86–88.
15. Goshtasby Arthur Ardeshir. *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005.
16. Huttenlocher D. P., Klanderman G. A., Rucklidge W. J. Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 1993, vol. 15, no. 9, pp. 850–863.
17. Болотова Ю. А., Спицын В. Г., Осина П. М. Обзор алгоритмов детектирования текстовых областей на изображениях и видеозаписях. *Компьютерная оптика*, 2017, Т. 41, № 3, стр. 441–452.
18. Трипутень В. В., Гороховатский В. А. Алгоритм параллельной нормализации аффинных преобразований для цветных изображений. *Радиоэлектроника и информатика*, 1997, № 1, стр. 97–98.
19. Егошкин Н. А., Еремеев В. В., Козлов Е. П. Нормализация космических изображений Земли на основе их сопоставления с электронными картами. *Цифровая обработка сигналов*, 2009, № 3, стр. 21–26.

20. Singh S. K., Naidu S. D., Srinivasan T. P., Krishna B. G., Srivastava P.K. Rational polynomial modelling for cartosat-1 data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2008, vol. 37, pp. 885–888.
21. Safari R., Narasimhamurthi N., Shridhar Malayappan and Ahmadi Majid. Document registration using projective geometry. *IEEE transactions on image processing*, 1997, vol. 6, no. 9, pp. 1337–1341.
22. Iwamura Masakazu, Niwa Ryo, Kise Koichi, Uchida Seiichi and Omachi Shinichiro. Rectifying perspective distortion into affine distortion using variants and invariants. *Proceedings of the Second International Workshop on Camera-Based Document Analysis and Recognition*, 2007, pp. 138–145.
23. 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.
24. Шемякина Ю. А., Жуковский А. Е., Фараджев И. А. Исследование алгоритмов вычисления проективного преобразования в задаче наведения на планарный объект по особым точкам *Искусственный интеллект и принятие решений*, 2017, Т. 2017, № 1, стр. 43–49.
25. Xie Youye, Tang Gongguo, Hoff William. Geometry-based populated chessboard recognition. *Tenth International Conference on Machine Vision (ICMV 2017)*, 2018, vol. 1069603, pp. 1–5.
26. Jesorsky O., Kirchberg K. J., Frischholz R. W. Robust face detection using the hausdorff distance. *International Conference on Audio-and Video-Based Biometric Person Authentication*. Berlin, Heidelberg: Springer, 2001, pp. 90–95.
27. Zeynalov Ramiz, Velizhev Alexander, Konushin Anton. Восстановление формы страницы текста для коррекции геометрических искажений. *Proc. of the 19 International Conference GraphiCon-2009*, Moscow, Russia, 2009, стр. 125–128.
28. Dubuisson M. P., Jain A. K. A modified Hausdorff distance for object matching. *Proceedings of 12th international conference on pattern recognition*. IEEE, 1994, vol. 1, pp. 566–568.
29. Orrite C., Herrero J. E. Shape matching of partially occluded curves invariant under projective transformation. *Computer Vision and Image Understanding*, 2004, vol. 93, no. 1, pp. 34–64.
30. Николаев П. П. Проективно инвариантное описание неплоских гладких фигур. 1. Предварительный анализ задачи. *Сенсорные системы*, 2016, Т. 30, № 4, стр. 290–311.
31. Skoryukina N., Chernov T., Bulatov K., Nikolaev D. P., Arlazarov V. TV-stream frame search with projectively distorted and noisy query. *Ninth International Conference on Machine Vision (ICMV 2016)*. SPIE, 2017, vol. 103410Y, pp. 1–5.
32. Arvind C. S., Mishra R., Vishal K., Gundimeda V. Vision based speed breaker detection for autonomous vehicle. *Tenth International Conference on Machine Vision (ICMV 2017)*. International Society for Optics and Photonics, 2018, vol. 106960E, pp. 1–9.
33. Ефимов А. И., Новиков А. И. Алгоритм поэтапного уточнения проективного преобразования для совмещения изображений. *Компьютерная оптика*, 2016, Т. 40, № 2, стр. 258–265.
34. Притула Н. Е., Николаев П. П., Шешкус А. В. Сравнение двух алгоритмов проективно-инвариантного распознавания плоских замкнутых контуров с единственной вогнутостью. *ITuS 2014*, 2014, стр. 367–373.
35. Awal A. M., Ghanmi N., Sicre R., Furon T. Complex document classification and localization application on identity document images. *IAPR 2017-International Conference on Document Analysis and Recognition*. IEEE, 2017, pp. 427–431.
36. 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.
37. Zhang Wangbo, Li Xiaoyuan and Ma Xingjie. Perspective correction method for Chinese document images. *Intelligent Information Technology Application Workshops, 2008. ITAW'08. International Symposium on*, 2008, pp. 467–470.

38. Arlazarov V. V., Slavin O. A. E., Uskov A. V. E., Janiszewski I. M. Modelling the flow of character recognition results in video stream. *Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming and Computer Software*, 2018, vol. 11, no. 2, pp. 14–28.
39. Legge G. E., Pelli D. G., Rubin G. S., Schleske M. M. Psychophysics of reading—I. Normal vision. *Vision research*, 1985, vol. 25, no. 2, pp. 239–252.
40. Karpenko S., Konovalenko I., Miller A., Miller B., Nikolaev D. UAV Control on the Basis of 3D Landmark Bearing-Only Observations. *Sensors*, 2015, vol. 15, no. 12, pp. 29802–29820.
41. Кунина И. А., Гладилин С. А., Николаев Д. П. Слепая компенсация радиальной дисторсии на одиночном изображении с использованием быстрого преобразования Хафа. *Компьютерная оптика*, 2016, Т. 40, № 3, стр. 395–403.
42. Tong L., Zhang Y. Correction of perspective text image based on gradient method. *2010 International Conference on Information, Networking and Automation (ICINA)*. IEEE, 2010, vol 2, pp. 312–316.
43. Takezawa Y., Hasegawa M., Tabbone S. Camera-captured document image perspective distortion correction using vanishing point detection based on Radon transform. *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 3968–3974.
44. Szeliski R. Video mosaics for virtual environments. *IEEE computer Graphics and Applications*, 1996, vol. 16, no. 2, pp. 22–30.
45. Sawhney H. S., Kumar R. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, vol. 21, no. 3, pp. 235–243.
46. Calderon F., Romero L. An accurate image registration method using a projective transformation model. *Eighth Mexican International Conference on Current Trends in Computer Science (ENC 2007)*. IEEE, 2007, pp. 58–64.
47. Гошин Е. В., Котов А. П., Фурсов В. А. Двухэтапное формирование пространственного преобразования для совмещения изображений. *Компьютерная оптика*, 2014, Т. 38, № 4, стр. 886–891.
48. Haker S., Zhu L., Tannenbaum A., Angenent S. Optimal mass transport for registration and warping. *International Journal of computer vision*, 2004, vol. 60, no. 3, pp. 225–240.
49. Schmitzer B., Schnörr C. Globally optimal joint image segmentation and shape matching based on Wasserstein modes. *Journal of Mathematical Imaging and Vision*, 2015, vol. 52, no. 3, pp. 436–458.
50. Su Zhengyu, Zeng Wei, Wang Yalin, Lu Zhong-Lin, Gu Xianfeng. Shape classification using Wasserstein distance for brain morphometry analysis. *International Conference on Information Processing in Medical Imaging*, 2015, pp. 411–423.
51. Jaccard Paul. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull Soc Vaudoise Sci Nat*, 1901, no. 37, pp. 241–272.
52. Zhukovskiy A. E., Nikolaev D. P., Arlazarov V. V., Postnikov V. V., Polevoy D. V., Skoryukina N. S., Chernov T. S., Shemyakina Y. A., Mukovozov A. A., Konovalenko I. A., Povolotskiy M. A. Segments Graph-Based Approach for Document Capture in a Smartphone Video Stream. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE Computer Society, 2018, vol. 1, pp. 337–342.
53. Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.
54. Sim D. G., Kwon O. K., Park R. H. Object matching algorithms using robust Hausdorff distance measures. *IEEE Transactions on image processing*, 1999, vol. 8, no. 3, pp. 425–429.
55. Pritula N., Nikolaev D. P., Sheshkus A., Pritula M., Nikolaev P. P. Comparison of Two Algorithms Modifications of Projective-Invariant Recognition of the Plane Boundaries with the One Concavity. *Seventh International Conference on Machine Vision (ICMV 2014)*. SPIE, 2015. vol. 944508 pp. 1–5.

56. Stein Gideon P. Lens distortion calibration using point correspondences. *Computer Vision and Pattern Recognition, 1997*, 1997, pp. 602–608.
57. Hsu S. C., Sawhney H. S. Influence of global constraints and lens distortion on pose and appearance recovery from a purely rotating camera. *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on.* IEEE, 1998. pp. 154–159.
58. Chen H., Sukthankar R., Wallace G., Li K. Scalable alignment of large-format multi-projector displays using camera homography trees. *Proceedings of the conference on Visualization'02.* IEEE Computer Society, 2002, pp. 339–346.
59. Катаманов С. Н. Метод проективного сопоставления для овалов с двумя отмеченными точками. *Современные проблемы дистанционного зондирования Земли из космоса*, 2007, Т. 1, № 4, стр. 63–68.
60. Dance C. R. Perspective estimation for document images. *Document Recognition and Retrieval IX.* International Society for Optics and Photonics, 2001, vol 4670, pp. 244–255.
61. Коноваленко И. А., Кохан В. В., Николаев Д. П. Оптимальная аффинная аппроксимация проективного преобразования изображений. *Сенсорные системы*, 2019, Т. 33, № 1, стр. 7–14, DOI: 10.1134/S0235009219010062.
62. Коноваленко И. А., Шемякина Ю. А. Анализ величин ошибки при неточном проективном преобразовании четырехугольника. *Сборник трудов ИТНТ-2018*, 2018, стр. 1251–1260.
63. Skoryukina N., Shemiakina J., Arlazarov V. L., Faradjev I. Document localization algorithms based on feature points and straight lines. *Tenth International Conference on Machine Vision (ICMV 2017).* SPIE, 2018, vol. 106961H, pp. 1–8.
64. Коноваленко И. А., Полевой Д. В., Николаев Д. П. Максимальная невязка направлений как критерий точности проективной нормализации изображения при оптическом распознавании текста. *Сенсорные системы*, 2020, Т. 34, № 2, стр. 131–146, DOI: 10.31857/S0235009220020079.
65. Zhang Z., He L. W. Whiteboard scanning and image enhancement. *Digital signal processing*, 2007, vol. 17, no. 2, pp. 414–432.
66. Forsyth David A. and Ponce Jean. Computer vision: a modern approach. *Prentice Hall Professional Technical Reference*, 2002.
67. Шемякина Ю. А. Использование точек и прямых для вычисления проективного преобразования по двум изображениям плоского объекта. *Информационные технологии и вычислительные системы*, vol. 2017, no. 3, pp. 79–91.
68. Hartley Richard and Zisserman Andrew. Multiple view geometry in computer vision. *Cambridge University Press*, 2003.

RMS coordinate discrepancy as accuracy criterion of images normalization at optical document recognition

Konovalenko I.A.

Abstract—Optical text recognition on an image made from an arbitrary angle requires preliminary normalization of the image – transformation to such a view as if it had been obtained from a convenient angle for recognition. In this case, inaccuracies in normalization lead to recognition errors. To date, a number of accuracy criteria of normalization have been proposed in the literature, but their compliance with the quality of recognition is not studied. In this paper, for the case of a document with a fixed structure, a normal probabilistic recognition model is introduced, according to which the probability of correct recognition of a character decreases along the Gaussian with an increase in the discrepancy of the coordinates of this character. For this model, it was proved that the the accuracy criterion of image normalization, which is equal to RMS (root mean square) coordinate discrepancy over the text fields of the document, is monotonically related to the probability of the correct recognition of the entire document. Finally, RMS coordinate discrepancy for important case of projective normalization and region of interest consisting of rectangles was calculated analytically for the first time.

KEYWORDS: image normalization, projective transformation, optical character recognition, document recognition, accuracy criteria, coordinate discrepancy.