

УДК 004.932.2

МАКСИМАЛЬНАЯ НЕВЯЗКА НАПРАВЛЕНИЙ КАК КРИТЕРИЙ ТОЧНОСТИ ПРОЕКТИВНОЙ НОРМАЛИЗАЦИИ ИЗОБРАЖЕНИЯ ПРИ ОПТИЧЕСКОМ РАСПОЗНАВАНИИ ТЕКСТА

© 2020 г. И. А. Коноваленко^{1,2,*}, Д. В. Полевой^{2,3,4,5}, Д. П. Николаев^{1,2}

¹ Институт проблем передачи информации РАН
127051 Москва, Большой Каретный переулок, д. 19, Россия

² ООО СмарТ Энджинс Сервис
117312 Москва, проспект 60-летия Октября, д. 9, Россия

³ Федеральный исследовательский центр “Информатика и управление” РАН
117312 Москва, проспект 60-летия Октября, д. 9, Россия

⁴ Национальный исследовательский технологический университет “МИСиС”
119049 Москва, Ленинский проспект, д. 4, Россия

⁵ Московский физико-технический институт
141701 г. Долгопрудный, Институтский переулок, д. 9, Россия

*E-mail: konovalenko@iitp.ru

Поступила в редакцию 08.10.2019 г.

После доработки 27.10.2019 г.

Принята к публикации 17.12.2019 г.

Общепринято применение проективной нормализации (частный случай ортокоррекции) к фотографиям документов для их дальнейшего оптического распознавания. На сегодняшний день известен ряд критериев точности проективной нормализации. Почти все они характеризуют только невязку координат. Однако текстовые поля документов обычно имеют вытянутую форму, так что даже при небольших невязках координат возможны большие невязки направлений, которые значительно влияют на качество сегментации поля и распознавания отдельных символов в нем. Проблема точной коррекции невязок направлений стоит и в задачах томографии, если для измерения используется спиральная схема сканирования или проекции регистрируются в схемах томосинтеза. Для описания точности проективной нормализации изображения в точке предлагается поточечная максимальная невязка направлений. В качестве критерия точности проективной нормализации всего изображения предлагается максимальная невязка направлений, равная максимальной по области интереса поточечной максимальной невязке направлений. Получено аналитическое решение задачи вычисления поточечной максимальной невязки направлений. Выдвинута и численно подтверждена гипотеза о том, что поточечная максимальная невязка направлений – квазивыпуклая функция. Доказана теорема о том, что супремум квазивыпуклой функции на ограниченном замкнутом множестве равен супремуму на крайних точках его выпуклой оболочки. На основании гипотезы и теоремы предложено аналитическое решение задачи вычисления максимальной невязки направлений на полиэдральной области интереса.

Ключевые слова: ортокоррекция, коррекция перспективы, проективная нормализация изображений, критерии точности, невязка направлений, оптическое распознавание символов, математическое программирование

DOI: 10.31857/S0235009220020079

ВВЕДЕНИЕ

Проективная нормализация изображений

Путем устранения радиальной дисторсии (Kunina et al., 2017) практически любая изображающая оптическая система может быть сведена к ортоскопической оптической системе, т.е. к такой изображающей оптической системе, которая описывается проективной моделью камеры (Ша-

пиро и др., 2013). В рамках проективной модели камеры произвольная плоскость сцены оказывается связана с плоскостью изображения проективным преобразованием.

Ракурс съемки относительно плоской сцены называется нормальным, если изображающая оптическая система (далее – камера) ориентирована по нормали к сцене. Имитация сделанного с виртуального нормального ракурса съемки изоб-

ражения на основе изображения, сделанного с произвольного ракурса, обеспечивается проективным преобразованием последнего. На основе работы (Путятин и др., 1998) будем называть такую имитацию проективной нормализацией изображения (ПНИ). Полученное в результате ПНИ изображение называется нормализованным. Нормализованное изображение является ортоскопическим, т.е. образ плоской сцены на нем подобен самой сцене, что принципиально упрощает ее дальнейший анализ. Как правило, при ПНИ от виртуального ракурса съемки требуют не только нормальности, но и фиксации тех или иных оставшихся степеней свободы камеры (твердого тела), которые проявляются как изотропное масштабирование, сдвиг и ориентация нормализованного изображения.

Проективная нормализация активно используется в качестве этапа предобработки изображений при решении различных задач технического зрения, таких как распознавание текстового содержания документов (Зейналов и др., 2009; Болотова и др., 2017; Шемякина и др., 2017; Skoryukina et al., 2018); распознавание автомобильных номеров (Povolotskiy et al., 2017); автоматическое узнавание телевизионной передачи по фотографии экрана телевизора (Skoryukina et al., 2017); детектирование шахматной доски с целью калибровки камеры (Xie Y et al., 2018); детектирование искусственной неровности на дорогах (Arvind et al., 2018); сопоставление контура представленного на изображении объекта с объектом в базе данных (Dubuisson, Jain, 1994; Sim et al., 1999; Orrite, Hergero, 2004; Николаев, 2016; Балицкий и др., 2017; Савчик, Николаев, 2018); томографическая реконструкция: при предобработке томографических проекций перед применением алгоритма реконструкции при использовании спиральной, конусной схемы сканирования (Бузмаков, 2019), спутниковый мониторинг (оценка временной изменчивости температуры поверхности океана, определение скорости передвижения облачных масс и многое другое) (Катаманов, 2007), составление планов и карт местности по результатам аэрофотосъемки (Karpenko et al., 2015; Холопов, 2017) и многих других. Кроме того, проективная нормализация фотографии документа применима для облегчения восприятия ее человеком (Legge et al., 1985).

Обзор геометрических критериев точности проективной нормализации изображений

Как было показано выше, ПНИ используется в качестве этапа предобработки изображений при решении многих задач технического зрения. Данный раздел посвящен обзору существующих в литературе геометрических критериев точности ПНИ.

Начнем с введения используемых далее определений и обозначений:

I_{input} – входное изображение (рис. 1). Обычно им является фотография.

H – проективное преобразование, задающее идеальную нормализацию изображения I_{input} . Далее будем рассматривать только случаи, когда H – единственное. На практике H не известно, но считается известным при тестировании методов ПНИ.

\hat{H} – проективное преобразование, получаемое методом ПНИ и оценивающее H .

I_{ideal} – идеально нормализованное изображение. Формируется в результате применения H к I_{input} .

I_{pract} – практически нормализованное изображение. Формируется в результате применения \hat{H} к I_{input} .

$\mathbf{x} = [x_1 \ x_2]^T$ – декартовы координаты пикселей на плоскости изображения I_{ideal} .

$\mathbf{y} = [y_1 \ y_2]^T$ – декартовы координаты пикселей на плоскости изображения I_{pract} .

$\mathbf{V} \stackrel{\text{def}}{=} \hat{H}H^{-1}$ – остаточное проективное преобразование, для каждой точки сцены переводящее координаты ее образа на изображении I_{ideal} в координаты ее образа на изображении I_{pract} : $\mathbf{y} = \mathbf{V}(\mathbf{x})$.

$d(\mathbf{x}|\mathbf{V}) \stackrel{\text{def}}{=} \|\mathbf{x} - \mathbf{V}(\mathbf{x})\|_2$ – невязка координат (Kunina et al., 2016), поточечно характеризующая ошибку ПНИ.

$R \subset \mathbb{R}^2$ – занятое образом целевого объекта сцены (документа, автомобильного номера, здания и другое) множество точек изображения I_{ideal} , называемое областью интереса.

$Q \stackrel{\text{def}}{=} \mathbf{V}[R] \stackrel{\text{def}}{=} \{\mathbf{V}(\mathbf{x}) : \mathbf{x} \in R\}$ – образ области интереса R на изображении I_{pract} .

Перейдем непосредственно к обзору геометрических критериев точности ПНИ:

1. На конкурсе “Smartphone document capture” конференции ICDAR (Zhukovsky et al., 2017) в качестве критерия выбран коэффициент Жаккарда (Jaccard, 1901), который равен площади пересечения множеств Q и R , отнесенной к площади их объединения:

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

2. В работе (Sim et al., 1999) для проективного сопоставления объектов используется метрика Хаусдорфа, равная наибольшему расстоянию от точек одного множества до соответствующих им ближай-

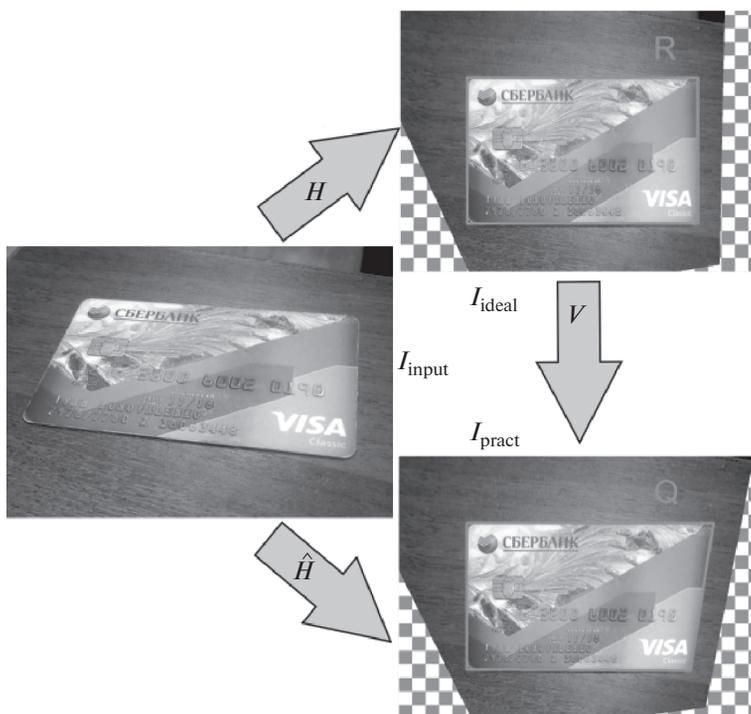


Рис. 1. Общая схема преобразований.

I_{input} – сделанная с произвольного ракурса фотография документа, I_{ideal} – идеально нормализованное изображение, I_{pract} – практически нормализованное изображение.

ших точек второго множества: $d_{Hausdorff}(Q, R) \stackrel{\text{def}}{=} \max \left\{ \max_{y \in Q} \min_{x \in R} \|y - x\|_2, \max_{x \in R} \min_{y \in Q} \|x - y\|_2 \right\}$.

В статье (Orrite, Herrero, 2004) она используется для проективного выравнивания частично замкнутых контуров, в статье (Jesorsky et al., 2001) – для робастной детекции лиц, в работе (Huttenlocher et al., 1993) – для вычисления близости двух изображений. В ряде работ опубликованы ее модификации (Dubuisson, Jain, 1994; Ефимов, Новиков, 2016).

3. В статье (Притула и др., 2014) для измерения близости двух контуров после проективного выравнивания в задаче проективно-инвариантного распознавания плоских замкнутых контуров применяется расстояние Фреше (Fréchet, 1906), которое определяется следующим образом:

$F(Q, R) = \inf_{a,b} \max_{t \in [0,1]} \|\delta Q(a(t)), \delta R(b(t))\|_2$, 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 используется также для вычисления точности сегментации изображе-

ний (Березский, Березская, 2015) и точности сопоставления карт (Wei et al., 2013).

4. В задаче автоматической коррекции дисторсии, вызванной кривизной линзы и движением камеры (Hsu, Sawhney, 1998), используется среднеквад-

ратичный критерий: $L_2(V; R) \stackrel{\text{def}}{=} \sqrt{\frac{1}{S(R)} \int_R d^2(x|V) dx}$,

где $S(R)$ – площадь области интереса R . Он же используется для вычисления точности формирования единого изображения при помощи матрицы проекторов (Chen et al., 2002).

5. В работе (Катаманов, 2007) для вычисления точности автоматической привязки, полученной с геостационарного спутника изображений, использовался минимаксный критерий, равный максимальному значению невязки координат на

области интереса R : $L_\infty(V; R) \stackrel{\text{def}}{=} \max_{x \in R} d(x|V)$.

6. Некоторые критерии предназначены только для случая, когда область интереса R является прямоугольником. Тогда ее искаженный образ Q является четырехугольником. Например, в работах (Rodríguez-Piñeiro et al., 2011; Zhang, He, 2007; Takezawa et al., 2016; Awal et al., 2017) для оценки точности ПНИ используется ошибка в соотношении сторон четырехугольника Q . В работе (Calore et al., 2012) для оценки точности устранения про-

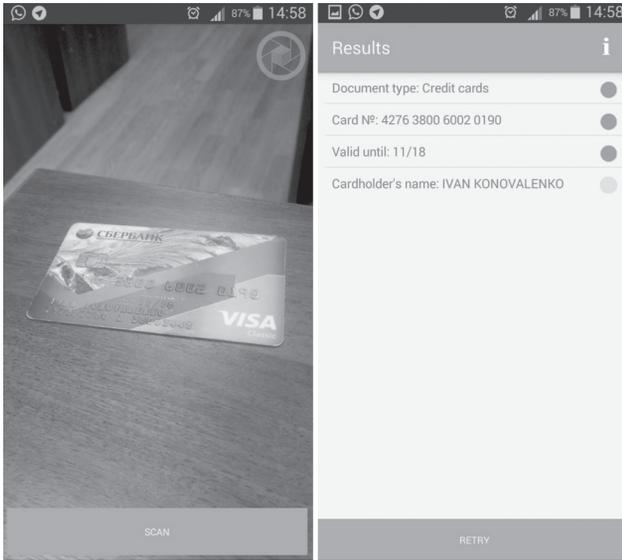


Рис. 2. Пример работы системы распознавания. Слева направо: фотография документа, распознанное текстовое содержание документа.

активных искажений используется угол между противоположными сторонами четырехугольника Q , а в работе (Холопов, 2017) – критерий, равный отношению минимального и максимального углов четырехугольника Q .

Заметим, что все найденные в литературе геометрические критерии точности зависят от \hat{N} и N только через $V = \hat{N}N^{-1}$. Поэтому точность оценки \hat{N} преобразования N можно понимать как близость остаточного преобразования V к тождественному преобразованию I .

ЗАДАЧА РАСПОЗНАВАНИЯ ДОКУМЕНТА ПО СДЕЛАННОЙ С ПРОИЗВОЛЬНОГО РАКУРСА ФОТОГРАФИИ

Рассмотрим следующую задачу компьютерного зрения. Пусть дана сделанная с произвольного ракурса фотография физически плоского документа. Требуется распознать текстовое содержание документа (Arlazarov et al., 2018). Для краткости в рамках данной работы назовем эту задачу “задачей распознавания”. Решающую задачу распознавания системы назовем “системой распознавания”. Пример работы системы распознавания приведен на рис. 2.

Будем считать, что система распознавания декомпозирована на следующие два модуля (рис. 3) (Zhukovsky et al., 2017; Шемякина и др., 2017; Skoyukina et al., 2018).

- Модуль ПНИ изображения, который на основе входного изображения I_{input} вычисляет оценку \hat{N} идеальной нормализации N , и, применяя ее к изображению I_{input} , формирует практически нормализованное изображение I_{pract} .

- Модуль распознавания, который принимает на вход практически нормализованное изображение I_{pract} и возвращает текстовое содержание документа.

Пользуясь вышеприведенными данными, проанализируем известные критерии точности ПНИ на предмет их соответствия качеству последующего решения задачи распознавания. Большинство из них характеризуют только невязку координат. Однако текстовые поля документов обычно имеют вытянутую форму, так что даже при небольших невязках координат возможны большие невязки направлений (рис. 4), которые значительно влияют на качество сегментации поля и



Рис. 3. Декомпозиция задачи распознавания.



Рис. 4. Практически нормализованные изображения I_{pract} : a – без искажения (I_{ideal}), b – с аффинным искажением, v – с проективным искажением.

распознавания отдельных символов в нем (Bezmaternykh et al., 2018). Критерии точности, характеризующие невязку направлений, также встречаются в литературе (пункт б), однако, их связь с качеством распознавания представляется недостаточной или, по крайней мере, недостаточно ясной. В следующем разделе мы предлагаем новый критерий на основе невязок направлений, специально построенный для описания точности ПНИ в задаче распознавания.

МАКСИМАЛЬНАЯ НЕВЯЗКА НАПРАВЛЕНИЙ ПРОЕКТИВНОЙ НОРМАЛИЗАЦИИ ИЗОБРАЖЕНИЯ

Угловой модуль

Для описания абсолютной разницы между углами α и β введем угловой модуль, равный углу между векторами $\begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix}$ и $\begin{bmatrix} \cos(\beta) \\ \sin(\beta) \end{bmatrix}$:

$$\langle \beta - \alpha \rangle \stackrel{\text{def}}{=} \arccos(\cos(\beta - \alpha)) = |((\beta - \alpha + \pi) \bmod 2\pi) - \pi|. \tag{1}$$

Угловой модуль отличается от обычного модуля тем, что учитывает периодичность углов (рис. 5).

Невязка направлений

Напомним, что остаточное проективное преобразование V для каждой точки сцены переводит координаты ее образа на изображении I_{ideal} в координаты ее образа на изображении I_{pract} : $\mathbf{u} = V(\mathbf{x})$. Сперва рассмотрим случай общего положения, который заключается в том, что матрица Якоби $J(\mathbf{x})$ преобразования V существует и невырождена в каждой точке выпуклой оболочки

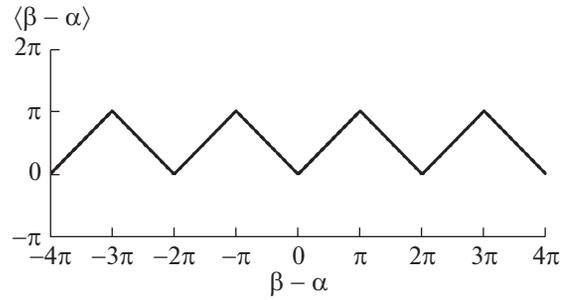


Рис. 5. Угловой модуль.

области интереса $\mathbf{x} \in \text{Conv}(R)$. Тогда бесконечно малый вектор, исходящий из точки $\mathbf{x} \in \text{Conv}(R)$ по направлению α , после преобразования V перейдет в бесконечно малый вектор, исходящий из точки $V(\mathbf{x})$ по направлению β , которое можно рассчитать по формуле

$$\beta(\mathbf{x}, \alpha | V) = \lim_{R \rightarrow +0} \arctg2 \left[V \left(\mathbf{x} + R \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix} \right) - V(\mathbf{x}) \right], \tag{2}$$

где $\arctg2[\mathbf{b}]$ – измеренная в радианах ориентация вектора \mathbf{b} . Отметим, что различные направления α поворачиваются на разные углы $\langle \beta - \alpha \rangle$. Осуществляя предельный переход, линейризуем аргумент арктангенса в точке \mathbf{x} , тогда

$$\beta(\mathbf{x}, \alpha | V) = \arctg2 \left[J(\mathbf{x}) \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix} \right]. \tag{3}$$

Проективное преобразование по определению сохраняет прямые, поэтому вместо бесконечно малых векторов с тем же результатом можно рассматривать произвольные. Таким образом, радиус R в формуле (2) можно не устремлять к нулю, а выбирать его произвольно положительным (рис. 6, з):

$$\beta(\mathbf{x}, \alpha | V) = \arctg2 \left[V \left(\mathbf{x} + R \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix} \right) - V(\mathbf{x}) \right], \tag{4}$$

$R > 0$.

Невязкой направлений назовем функцию, которая в каждой точке $\mathbf{x} \in \text{Conv}(R)$ равняется искажению каждого направления α остаточным преобразованием V :

$$\Delta(\mathbf{x}, \alpha | V) \stackrel{\text{def}}{=} \langle \beta(\mathbf{x}, \alpha | V) - \alpha \rangle.$$

Примеры графиков Δ для различных \mathbf{x} и V можно видеть на рис. 6, 7 и 8. Так как проективное преобразование сохраняет прямые, невязка направлений, как функция от \mathbf{x} , не меняется по направлению α . В силу дифференцируемости проективного преобразования невязка направлений имеет период, равный π .

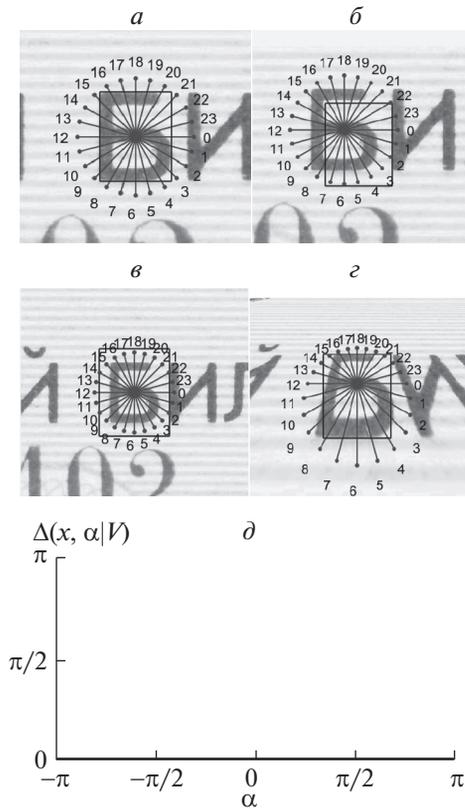


Рис. 6. Геометрические искажения изображений, не дающие невязок направлений во всех точках: *a* – тождественное преобразование; *б* – сдвиг; *в* – изотропное масштабирование; *г* – проективное преобразование (не дает невязок направлений в одной точке); рамками показано идеальное положение изображения символа; *д* – соответствующий вышеуказанным случаям график невязки направлений.

Поточечная максимальная невязка направлений

Известно, что невязки направлений различных типов (скос, поворот) создают сложности при сегментации и распознавании текста (Vezmaternykh et al., 2018). При этом неизвестно случая, чтобы большая невязка по какому-то из направлений не создавала таковых сложностей. Более того, невязки направлений $\Delta(x, \alpha|V)$ – функции не произвольные, а всего лишь трехпараметрические – по числу элементов матрицы Якоби (3) без учета масштаба, и большое значение невязки по одному направлению влечет за собой сопоставимые значения по большей части оставшихся направлений. Таким образом, для возникновения сложностей при сегментации и распознавании текста достаточно большой невязки хотя бы по одному из направлений. Поэтому в качестве критерия точности ПНИ в точке мы предлагаем поточечную максимальную невязку направлений:

$$\Delta_{\infty}(x|V) = \max_{\alpha} \Delta(x, \alpha|V). \tag{4}$$

Примеры графиков Δ_{∞} для различных V можно видеть на рис. 9 и 10.

Максимальная невязка направлений

В качестве скалярного интегрального критерия точности проективной нормализации изображения I_{pract} предлагается максимальная невязка направлений, равная максимальной по области интереса R поточечной максимальной невязке направлений:

$$\Delta_{\infty}^{\infty}(V; R) = \max_{x \in R} \Delta_{\infty}(x|V) = \max_{x \in R} \max_{\alpha} \Delta(x, \alpha|V). \tag{5}$$

Остаточное преобразование V – проективное, а значит, преобразует прямые в прямые. Следовательно, критерий Δ_{∞}^{∞} равняется наибольшему среди углов, на которые преобразование V поворачивает всевозможные прямые l , пересекающие область интереса R :

$$\Delta_{\infty}^{\infty}(V; R) = \max_{l \cap R \neq \emptyset} \Delta(l|V).$$

Таким образом, мы построили критерий точности ПНИ, “максимально” чувствительный к невязкам направлений. Напомним, что мотивацией для построения такого критерия послужила недостаточная чувствительность к невязкам направлений критериев, основанных на невязках координат. Однако критерий Δ_{∞}^{∞} сам по себе тем более не может претендовать на связь с качеством последующего распознавания нормализованного изображения, так как равняется строго нулю при сколь угодно большом сдвиге и изотропном масштабировании (рис. 6). Решение этой проблемы авторы видят в комбинированном использовании критерия максимальной невязки направлений Δ_{∞}^{∞} и критерия максимальной невязки координат

$$L_{\infty}(V; R) = \max_{x \in R} d(x|V),$$

ранее исследованного в работе (Konovalenko, Shemiakina, 2018).

Вырожденные случаи

Выше мы рассматривали только случай общего положения, когда матрица Якоби $J(x)$ преобразования V существует и невырождена в каждой точке x выпуклой оболочки области интереса $Conv(R)$. В этом разделе рассмотрим остальные случаи, которые будем называть вырожденными.

• Пусть существуют точки $Conv(R)$, в которых не существует матрицы Якоби. Выясним, что это за точки. Параметризуем преобразование V однородной матрицей $V = (v_{ij}) \in \mathbb{R}^{3 \times 3}$ следующим образом:

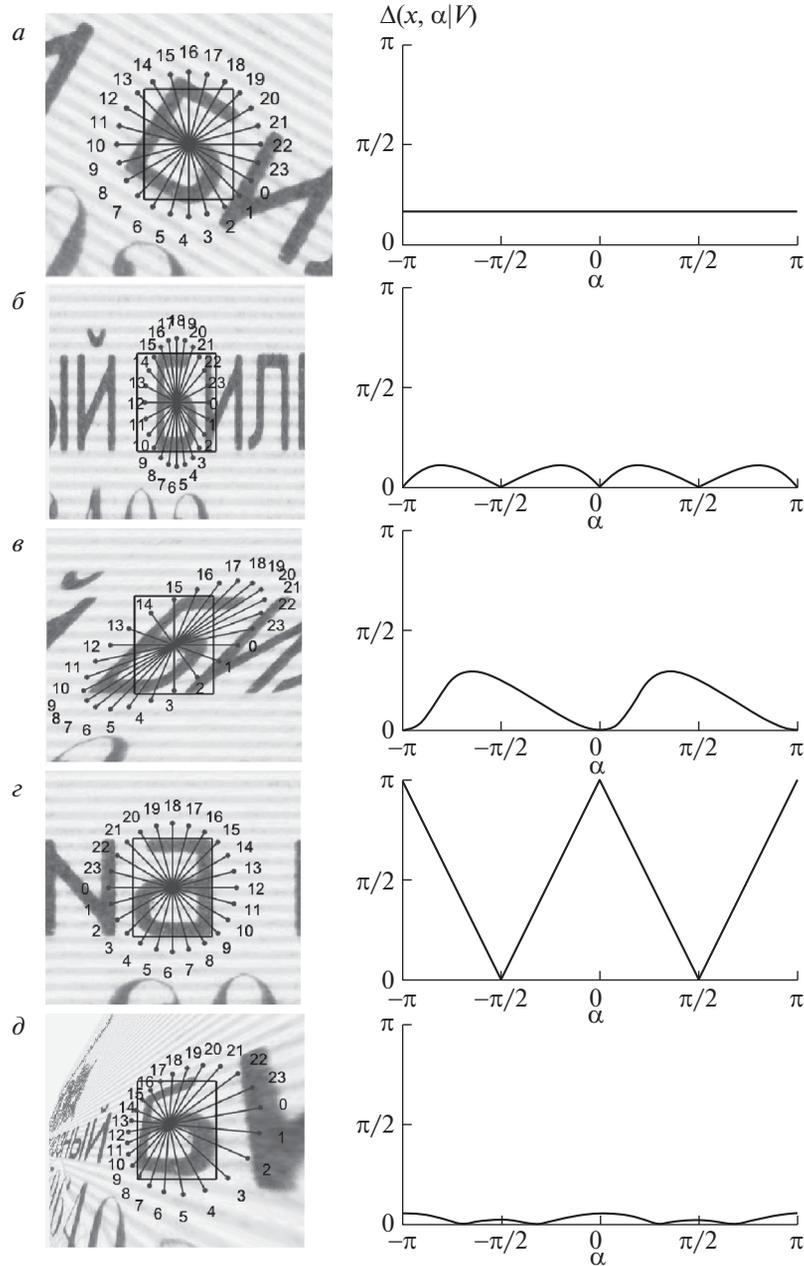


Рис. 7. Геометрические искажения изображений (слева) и соответствующие графики невязки направлений (справа): *a* – поворот; *б* – анизотропное масштабирование; *в* – скос; *г* – отражение; *д* – проективное преобразование.

$$V(x) \stackrel{\text{def}}{=} \frac{\begin{bmatrix} v_{11}x_1 + v_{12}x_2 + v_{13} \\ v_{21}x_1 + v_{22}x_2 + v_{23} \end{bmatrix}}{v_{31}x_1 + v_{32}x_2 + v_{33}}. \quad (6)$$

Прямую с уравнением $v_{31}x_1 + v_{32}x_2 + v_{33} = 0$, на которой знаменатель преобразования V обращается в ноль, будем называть горизонтом и обозначать $I_\infty \subset \mathbb{R}^2$. Легко видеть, что матрица Якоби преобразования V существует во всех точках евклидовой плоскости за исключением горизонта: $\mathbb{R}^2 \setminus I_\infty$. Соответственно поточечная максимальная

невязка направлений Δ_∞ для $x \in I_\infty$ формулой (4) не определяется. Однако оказывается, что Δ_∞ равняется π всюду по одну из сторон от горизонта. Это объясняется тем, что по “ту” сторону горизонта происходит отражение изображения, а значит, всегда найдется направление, противоположное исходному. Поэтому поточечная максимальная невязка направлений естественно доопределяется:

$$x \in I_\infty \Rightarrow \Delta_\infty(x|V) \stackrel{\text{def}}{=} \pi. \quad (7)$$

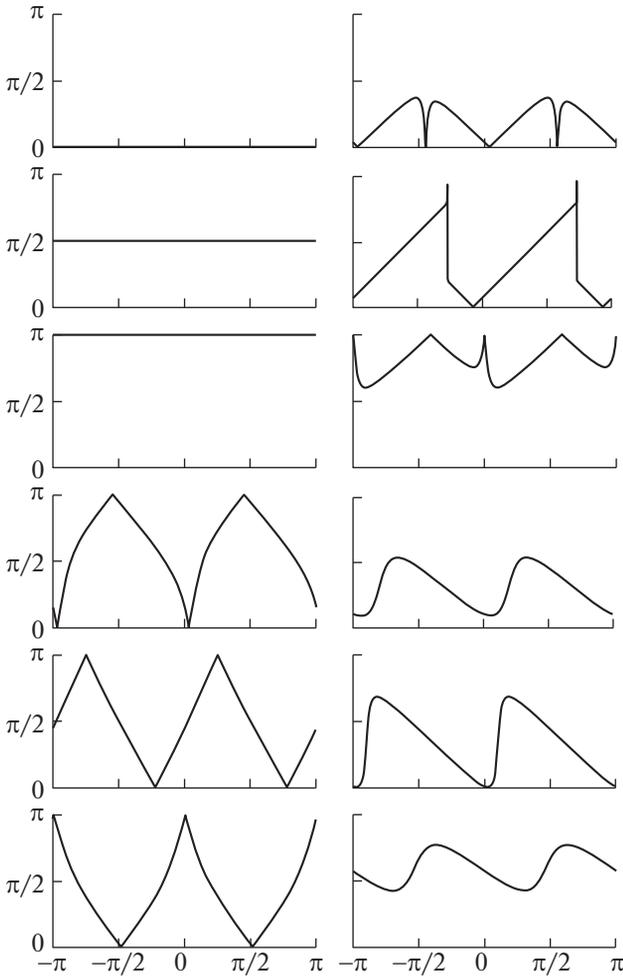


Рис. 8. Разнообразие невязок направлений $\Delta(x, \alpha|V)$.

• Пусть существуют точки $\text{Con}v(R)$, в которых матрица Якоби вырождена. Для проективного преобразования, как и для линейного, вырожденность матрицы Якоби в одной точке влечет ее вырожденность во всех точках. При этом плоскость изображения I_{ideal} схлопывается в прямую или точку. Поточечная максимальная невязка направлений Δ_∞ в таком случае формулой (4) не определяется и смысла не имеет.

ВЫЧИСЛЕНИЕ МАКСИМАЛЬНОЙ НЕВЯЗКИ НАПРАВЛЕНИЙ ПРОЕКТИВНОЙ НОРМАЛИЗАЦИИ ИЗОБРАЖЕНИЯ

В данном разделе мы построим аналитическое решение задачи вычисления максимальной невязки направлений (5) на полиэдре.

Вычисление поточечной максимальной невязки направлений

Начнем с подзадачи вычисления поточечной максимальной невязки направлений (4):

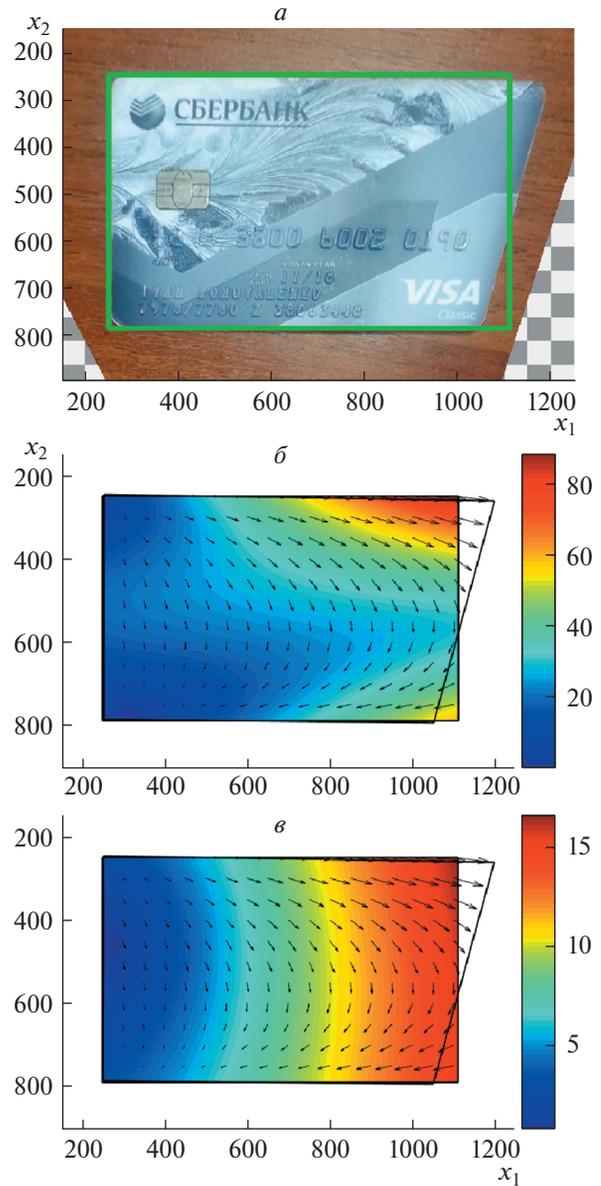


Рис. 9. Невязки остаточного преобразования V : a – практически нормализованное изображение I_{pract} ; зеленая рамка ограничивает идеальное положение образа документа; b – невязка координат $d(x|V) = \|x - V(x)\|_2$, измеренная в пикселях; c – поточечная максимальная невязка направлений $\Delta_\infty(x|V)$, измеренная в градусах; цвет соответствует значениям невязок; стрелочками показано векторное поле преобразования $V(x) - x$.

$$\Delta_\infty(x|V) = \max_{\alpha} \Delta(x, \alpha|V).$$

Если $x \in I_\infty$, то $\Delta_\infty(x|V) = \pi$ по формуле (7). Далее рассмотрим случай $x \notin I_\infty$. Запишем задачу с использованием формулы (3):

$$\Delta_\infty(x|V) = \max_{\alpha} \left\langle \left(\arctg 2 \left[J(x) \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix} \right] \right) - \alpha \right\rangle. \quad (8)$$

Матрица Якоби – константа, обозначим ее $J \stackrel{\text{def}}{=} J(\mathbf{x})$. Переменный единичный вектор обозначим $\mathbf{a}(\alpha) \stackrel{\text{def}}{=} \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix}$. Кроме того, обозначим $\mathbf{b}(\mathbf{a}) \stackrel{\text{def}}{=} J\mathbf{a}$ и $\beta(\mathbf{b}) \stackrel{\text{def}}{=} \arctg 2[\mathbf{b}]$, тогда $\Delta_\infty(\mathbf{x}|V) = \max_\alpha \langle \beta(\mathbf{b}(\mathbf{a}(\alpha))) - \alpha \rangle$.

Функция $\langle \beta(\mathbf{b}(\mathbf{a}(\alpha))) - \alpha \rangle$ может достигать максимума только в двух случаях: если недифференцируемая функция $\langle \bullet \rangle$ (1) равна своему локальному и глобальному максимуму π и если равняется нулю производная ее аргумента. Найдем α , соответствующие каждому из этих случаев.

- Пусть $\langle \beta(\mathbf{b}(\mathbf{a}(\alpha))) - \alpha \rangle = \pi$. Тогда в соответствии с определением $\langle \bullet \rangle$ или рис. 5: $\beta(\mathbf{b}(\mathbf{a}(\alpha))) - \alpha = \pi + 2\pi n$, $n \in \mathbb{Z}$. Значит, вектор $\mathbf{b} = J\mathbf{a}$ противоположен вектору \mathbf{a} : $J\mathbf{a} = \lambda\mathbf{a}$, $\lambda < 0$. В нашем случае \mathbf{a} – это собственные векторы матрицы J , соответствующие ее вещественным отрицательным собственным числам λ . Тогда искомые $\alpha = \arctg 2[\mathbf{a}]$ – ориентации таких векторов.

- Пусть

$$(\beta(\mathbf{b}(\mathbf{a}(\alpha))) - \alpha)'_\alpha = 0 \Rightarrow \beta'_b \mathbf{b}'_a \mathbf{a}'_\alpha = 1. \quad (9)$$

Возьмем производные:

$$\begin{aligned} \beta'_b &= (\mathbf{b}^T \mathbf{b})^{-1} \mathbf{b}^T M^T, \\ \mathbf{b}'_a &= J, \\ \mathbf{a}'_\alpha &= M\mathbf{a}, \end{aligned}$$

где

$$M \stackrel{\text{def}}{=} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

поэтому (9) опишем так:

$$\begin{aligned} (\mathbf{b}^T \mathbf{b})^{-1} \mathbf{b}^T M^T J M \mathbf{a} &= 1, \\ \mathbf{b}^T M^T J M \mathbf{a} &= \mathbf{b}^T \mathbf{b}, \\ \mathbf{a}^T J^T M^T J M \mathbf{a} &= \mathbf{a}^T J^T J \mathbf{a}, \\ \mathbf{a}^T (J^T M^T J M - J^T J) \mathbf{a} &= 0. \end{aligned}$$

Обозначим $K \stackrel{\text{def}}{=} (k_{ij}) \stackrel{\text{def}}{=} J^T M^T J M - J^T J$, тогда

$$\mathbf{a}^T K \mathbf{a} = 0,$$

$$\begin{aligned} k_{11} \cos^2(\alpha) + (k_{12} + k_{21}) \cos(\alpha) \sin(\alpha) + k_{22} \sin^2(\alpha) &= 0, \\ k_{22} \text{tg}^2(\alpha) + (k_{12} + k_{21}) \text{tg}(\alpha) + k_{11} &= 0. \end{aligned}$$

Решая это квадратное уравнение в вещественных числах, находим значения $\text{tg}(\alpha)$, из которых получаем искомые α .

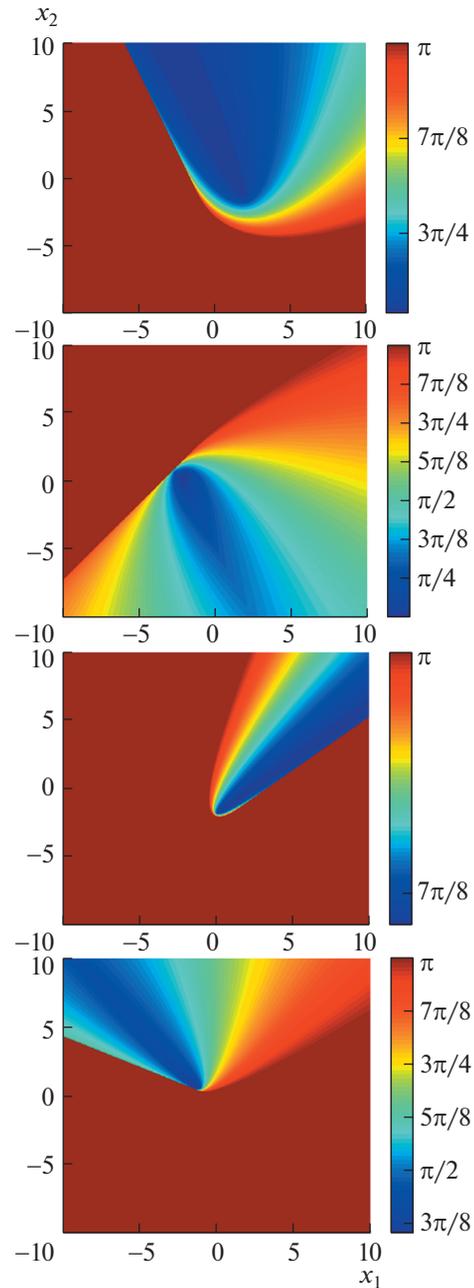


Рис. 10. Разнообразие поточечных максимальных невязок направлений $\Delta_\infty(\mathbf{x}|V)$.

Тогда искомый максимум (8) достигается на одном из значений α , вычисленных выше. Таким образом, получено аналитическое решение задачи вычисления поточечной максимальной невязки направлений Δ_∞ . Сформулируем алгоритм вычисления поточечной максимальной невязки направлений.

Вход: Точка $\mathbf{x} = [x_1 \ x_2]^T$ изображения I_{ideal} и однородная матрица $(v_{ij}) \in \mathbb{R}^{3 \times 3}$, определяющая (6) остаточное проективное преобразование V .

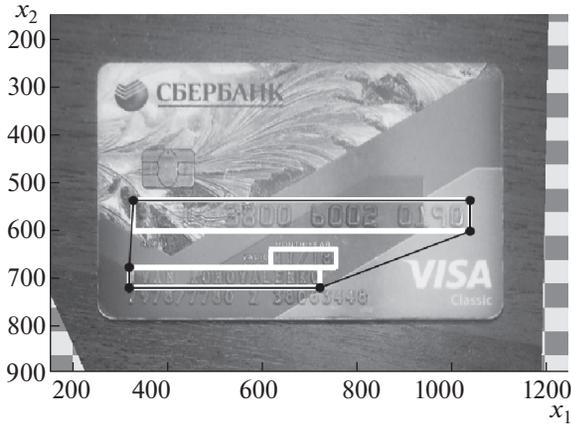


Рис. 11. Вычисление максимальной невязки направлений: белыми контурами ограничена полиэдральная область интереса R ; черный контур – многоугольник $\text{Conv}(R)$; черные точки – его вершины $E(\text{Conv}(R))$, на которых достигается максимальная невязка направлений $\Delta_\infty^\infty(V; R)$.

Выход: поточечная максимальная невязка направлений $\Delta_\infty(\mathbf{x}|V)$.

– Если $v_{31}x_1 + v_{32}x_2 + v_{33} = 0$, то $\Delta_\infty(\mathbf{x}|V) = \pi$. Иначе далее.

– Рассчитаем матрицу Якоби J проективного преобразования V в точке \mathbf{x} .

– Найдем все собственные векторы $\{\mathbf{a}_j\}_{j=1}^m$ матрицы J , соответствующие ее вещественным отрицательным собственным числам.

– Рассчитаем углы $\alpha_j = \arctg2[\mathbf{a}_j]$.

– Рассчитаем матрицу $(k_{ij}) = J^T M^T J M - J^T J$,

где $M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

– Найдем все вещественные корни $\{t_i\}_{i=1}^n$ квадратного уравнения

$$k_{22}t^2 + (k_{12} + k_{21})t + k_{11} = 0.$$

– Рассчитаем углы $\alpha'_i = \arctg[t_i]$.

– Сформируем набор углов $A = \{\{\alpha_j\}_{j=1}^m, \{\alpha'_i\}_{i=1}^n\}$.

– Рассчитаем поточечную максимальную невязку направлений:

$$\Delta_\infty(\mathbf{x}|V) = \max_{\alpha \in A} \Delta(\mathbf{x}, \alpha|V).$$

Вычисление максимальной невязки направлений

Теперь перейдем непосредственно к задаче вычисления максимальной невязки направлений (5): $\Delta_\infty^\infty(V; R) = \max_{\mathbf{x} \in R} \Delta_\infty(\mathbf{x}|V)$.

Проведенные авторами численные эксперименты позволяют выдвинуть гипотезу о том, что поточечная максимальная невязка направлений $\Delta_\infty(\mathbf{x}|V)$ проективного преобразования – квази-выпуклая функция (рис. 9 и 10). Примем эту гипотезу и допустим, что область интереса R – ограниченное и замкнутое множество. Тогда, в соответствии с теоремой 2 (Приложение), супремум $\Delta_\infty(\mathbf{x}|V)$ на R равен супремуму на крайних точках выпуклой оболочки R :

$$\sup_{\mathbf{x} \in R} \Delta_\infty(\mathbf{x}|V) = \sup_{\mathbf{x} \in E(\text{Conv}(R))} \Delta_\infty(\mathbf{x}|V).$$

Легко видеть, что функция $\Delta_\infty(\mathbf{x}|V)$ непрерывна, поэтому по теореме Вейерштрасса имеем $\max_{\mathbf{x} \in R} \Delta_\infty(\mathbf{x}|V) = \max_{\mathbf{x} \in E(\text{Conv}(R))} \Delta_\infty(\mathbf{x}|V)$, тогда

$$\Delta_\infty^\infty(V; R) = \max_{\mathbf{x} \in E(\text{Conv}(R))} \Delta_\infty(\mathbf{x}|V),$$

что во многих случаях существенно упрощает вычисление Δ_∞^∞ . Рассмотрим основной для нашего практического приложения частный случай, когда область интереса R состоит из многоугольников, т.е. является полиэдром. Выпуклая оболочка полиэдра – выпуклый многогранник, а в нашем плоском случае – выпуклый многоугольник. Крайние точки выпуклого многоугольника – его вершины. Значит, максимальную невязку направлений достаточно искать среди вершин $\{\mathbf{x}_i\}_{i=1}^n$ многоугольника, являющегося выпуклой оболочкой R (рис. 11):

$$\Delta_\infty^\infty(V; R) = \max_i \Delta_\infty(\mathbf{x}_i|V). \quad (10)$$

Таким образом, имеем следующий алгоритм вычисления максимальной невязки направлений на полиэдре.

Вход: остаточное проективное преобразование V , полиэдральная область интереса R .

Выход: максимальная невязка направлений $\Delta_\infty^\infty(V; R)$.

– Найдем вершины многоугольника, являющегося выпуклой оболочкой R : $\{\mathbf{x}_i\}_{i=1}^n = E(\text{Conv}(R))$.

– Используя алгоритм вычисления $\Delta_\infty(\mathbf{x}|V)$, вычислим максимальную невязку направлений:

$$\Delta_\infty^\infty(V; R) = \max_i \Delta_\infty(\mathbf{x}_i|V).$$

ЗАКЛЮЧЕНИЕ

В работе приведен ряд известных в литературе геометрических критериев точности проективной нормализации изображений (ПНИ) и проанализирована целесообразность их использования для оценки качества последующего распознавания текста на нормализованном изображении. Показано, что известные критерии имеют недостаточную чувствительность к невяз-

кам направлений, возникающих вследствие не- точности ПНИ, тогда как таковые невязки оказы- вают сильное негативное воздействие на качество сегментации поля и распознавания отдельных символов в нем. Для устранения этого недостатка построен вспомогательный критерий точности ПНИ, “максимально” чувствительный именно к невязкам направлений: для описания точности ПНИ в точке выбрана поточечная максимальная невязка направлений, в качестве критерия точно- сти проективной нормализации всего изображе- ния предложена максимальная невязка направле- ний, равная максимальной по области интереса поточечной максимальной невязке направлений. Получено аналитическое решение задачи вычис- ления поточечной максимальной невязки на- правлений. Выдвинута и численно подтверждена гипотеза о том, что поточечная максимальная не- вязка направлений – квазивыпуклая функция. Доказана теорема о том, что супремум квазивы- пуклой функции на ограниченном замкнутом множестве равен супремуму на крайних точках его выпуклой оболочки. На основании гипотезы и теоремы предложено аналитическое решение задачи вычисления максимальной невязки на- правлений на полиэдральной области интереса.

Полученные результаты могут быть использо- ваны как для построения методов ПНИ, так и для дальнейшего развития критериев точности ПНИ.

Работа выполнена при частичной финансовой поддержке РФФИ (проекты № 18-29-26022, 18- 29-26035).

ПРИЛОЖЕНИЕ

О МАКСИМУМЕ КВАЗИВЫПУКЛОЙ ФУНКЦИИ НА ОГРАНИЧЕННОМ ЗАМКНУТОМ МНОЖЕСТВЕ

Теорема 1. Пусть $X \subseteq \mathbb{R}^n$ – произвольное не- пустое множество, f – квазивыпуклая функция, определенная в том числе на $\text{Conv}(X)$:

$$\text{dom } f \supseteq \text{Conv}(X). \tag{11}$$

Тогда

$$\sup_X f = \sup_{\text{Conv}(X)} f. \tag{12}$$

Доказательство. Введем обозначение супре- мума f на X :

$$s \stackrel{\text{def}}{=} \sup_X f, \tag{13}$$

тогда

$$f(\mathbf{x} \in X) \leq s. \tag{14}$$

Пользуясь утверждением (11), рассмотрим все точки $\text{Conv}(X)$, значение f в которых не превышает супремума f на X :

$$X_s \stackrel{\text{def}}{=} \{\mathbf{x} \in \text{Conv}(X): f(\mathbf{x}) \leq s\}, \tag{15}$$

тогда

$$X_s \subseteq \text{Conv}(X) \tag{16}$$

и

$$f(\mathbf{x} \in X_s) \leq s. \tag{17}$$

Из $X \subseteq \text{Conv}(X)$ и формулы (14) следует, что X удовлетворяет всем ограничениям определения X_s (15), значит,

$$X \subseteq X_s. \tag{18}$$

Однако из квазивыпуклости f следует, что X_s – выпуклое. Таким образом, X_s – выпуклое множе- ство, содержащее X . Тогда $\text{Conv}(X) \subseteq X_s$, как наименьшее выпуклое множество, содержащее X . Отсюда, с учетом формулы (16), имеем

$$X_s = \text{Conv}(X). \tag{19}$$

Из формул (13) и (18) следует $\sup_{X_s} f \geq s$, откуда, с учетом формулы (17), следует

$$\sup_{X_s} f = s. \tag{20}$$

Из формул (13), (19) и (20) получаем утвержде- ние (12), что требовалось доказать:

$$\sup_X f = \sup_{\text{Conv}(X)} f. \tag{21}$$

Теорема 1 может применяться в обе стороны: для замены X на часто более простое множество $\text{Conv}(X)$; для замены $\text{Conv}(X)$ на часто суще- ственно меньшее множество X .

Перед формулировкой следующей теоремы приведем определение: крайней точкой выпукло- го множества C в вещественном векторном про- странстве называется точка, не являющаяся сере- диной отрезка в C .

Теорема 2. Пусть $X \subseteq \mathbb{R}^n$ – непустое ограни- ченное замкнутое множество, f – квазивыпуклая функция, определенная в том числе на $\text{Conv}(X)$:

$$\text{dom } f \supseteq \text{Conv}(X).$$

Тогда

$$\sup_X f = \sup_{E(\text{Conv}(X))} f, \tag{22}$$

где $E(C)$ – множество крайних точек выпуклого множества C .

Доказательство. Обозначим

$$C \stackrel{\text{def}}{=} \text{Conv}(X). \tag{23}$$

Из того, что множество X – ограниченное и замкнутое, следует, что множество C – ограниченное, замкнутое и выпуклое (теорема Каратеодори) (Юдин, 2010). Тогда следствие 18.5.1 из монографии (Рокафеллар, 1973) утверждает, что множество C есть выпуклая оболочка своих крайних точек:

$$C = \text{Conv}(E(C)). \quad (24)$$

Подставляя формулу (23) в формулу (24), получаем

$$\begin{aligned} \text{Conv}(X) &= \text{Conv}(E(\text{Conv}(X))) \Rightarrow \\ &\Rightarrow \sup_{\text{Conv}(X)} f = \sup_{\text{Conv}(E(\text{Conv}(X)))} f. \end{aligned} \quad (25)$$

Применяя теорему 1 к обоим частям равенства (25), получаем утверждение (22), что требуется доказать.

Замечания: теорема 2 близка к утверждениям параграфа “Максимумы выпуклых функций” монографии (Рокафеллар, 1973), но, в отличие от них, не требует выпуклости функции f и множества X ; в условиях теоремы 2 от множества X не требуется не только выпуклости, но и связности.

СПИСОК ЛИТЕРАТУРЫ

- Балицкий А.М., Савчик А.В., Гафаров Р.Ф., Коноваленко И.А. О проективно инвариантных точках овала с выделенной внешней прямой. *Проблемы передачи информации*. 2017. Т. 53. № 3. С. 84–89.
- Березский О.Н., Березская К.М. Количественная оценка качества сегментации изображений на основе метрик. *Управляющие системы и машины*. 2015. № 6. С. 59–65.
- Болотова Ю.А., Спицын В.Г., Осина П.М. Обзор алгоритмов детектирования текстовых областей на изображениях и видеозаписях. *Компьютерная оптика*. 2017. Т. 41. № 3. С. 441–452.
- Бузмаков А.В., Асадчиков В.Е., Золотов Д.А., Чукалина М.В., Ингачева А.С., Кривоносов Ю.С. Лабораторные рентгеновские микрофотографы: методы предобработки экспериментальных данных. *Известия РАН. Серия физическая*. 2019. Т. 83 (2). С. 194–197.
<https://doi.org/10.1134/S0367676519020066>
- Ефимов А.И., Новиков А.И. Алгоритм поэтапного уточнения проективного преобразования для совмещения изображений. *Компьютерная оптика*. 2016. Т. 40. № 2. С. 258–265.
<https://doi.org/10.18287/2412-6179-2016-40-2-258-265>
- Зейналов Р., Велижев А., Конушин А. Восстановление формы страницы текста для коррекции геометрических искажений. *Международная конференция Графикон. М.*: 2009. С. 125–128.
- Катаманов С.Н. Автоматическая привязка изображений геостационарного спутника MTSAT-1R. *Современные проблемы дистанционного зондирования Земли из космоса*. 2007. Т. 1. № 4. С. 63–68.
- Николаев П.П. Проективно инвариантное описание неплоских гладких фигур. 1. Предварительный анализ задачи. *Сенсорные системы*. 2016. Т. 30. № 4. С. 290–311.
- Притула Н.Е., Николаев П.П., Шешкус А.В. Сравнение двух алгоритмов проективно-инвариантного распознавания плоских замкнутых контуров с единственной вогнутостью. *Сборник трудов ИТИС-14*. 2014. С. 367–373.
- Путятин Е.П., Прокопенко Д.О., Печеная Е.М. Вопросы нормализации изображений при проективных преобразованиях. *Радиоэлектроника и информатика*. 1998. № 2 (3). С. 82–86.
- Рокафеллар Р. *Выпуклый анализ*. М.: Мир, 1973. Т. 472. С. 34–39.
- Савчик А.В., Николаев П.П. Метод проективного сопоставления для овалов с двумя отмеченными точками. *Информационные технологии и вычислительные системы*. 2018. № 1. С. 60–67.
- Холопов И.С. Алгоритм коррекции проективных искажений при маловысотной съемке. *Компьютерная оптика*. 2017. Т. 41. № 2. С. 284–290.
- Шапиро Л., Стокман Д., Богуславский А.А., Соколов С.М. *Компьютерное зрение*. М.: БИНОМ, 2013. 763 с.
- 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. № 1. С. 43–49.
- Юдин Д.Б. *Математические методы управления в условиях неполной информации*. Издательская группа URSS, М. 2010. 400 с.
- 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. V. 11. № 2. P. 14–28.
<https://doi.org/10.14529/mmp180202>
- Arvind C.S., Mishra R., Vishal K., Gundimeda V. Vision based speed breaker detection for autonomous vehicle. *Tenth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2018. V. 106960E. P. 1–9.
<https://doi.org/10.1117/12.2311315>
- Awal A.M., Ghanmi N., Sicre R., Furon T. Complex document classification and localization application on identity document images. *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE. 2017. V. 1. P. 426–431.
<https://doi.org/10.1109/ICDAR.2017.77>
- Bezmaternykh P.V., Nikolaev D.P., Arlazarov V.L. Textual blocks rectification method based on fast Hough transform analysis in identity documents recognition. *Tenth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2018. V. 1069606. P. 1–6.
<https://doi.org/10.1117/12.2310162>
- Calore E., Pedersini F., Frosio I. Accelerometer based horizon and keystone perspective correction. *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. 2012. P. 205–209.
<https://doi.org/10.1109/I2MTC.2012.6229434>

- 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. P. 339–346.
- Dubuisson M.P., Jain A.K. A modified Hausdorff distance for object matching. *Proceedings of 12th international conference on pattern recognition. IEEE*. 1994. V. 1. P. 566–568.
<https://doi.org/10.1109/ICPR.1994.576361>.
- Fréchet M.M. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884–1940)*. 1906. V. 22. № 1. P. 1–72.
- Hsu S.C., Sawhney H.S. Influence of global constraints and lens distortion on pose and appearance recovery from a purely rotating camera. *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No. 98EX201)*. 1998. P. 154–159.
<https://doi.org/10.1109/ACV.1998.732873>.
- Huttenlocher D.P., Klanderman G.A., Rucklidge W.J. Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*. 1993. V. 15. № 9. P. 850–863.
<https://doi.org/10.1109/34.232073>
- Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*. 1901. V. 37. P. 547–579.
- 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*. 2001. P. 90–95.
- Karpenko S., Konovalenko I., Miller A., Miller B., Nikolaev D. UAV control on the basis of 3D landmark bearing-only observations. *Sensors*. 2015. V. 15. № 12. P. 29802–29820.
<https://doi.org/10.3390/s151229768>
- Konovalenko I.A., Shemiakina J.A. Error values analysis for inaccurate projective transformation of a quadrangle. *Journal of Physics: Conference Series. – IOP Publishing*. 2018. V. 1096. № 1. P. 1–9.
<https://doi.org/10.1088/1742-6596/1096/1/012038>
- Kunina I.A., Gladilin S.A., Nikolaev D.P. Blind radial distortion compensation in a single image using fast Hough transform. *Computer optics*. 2016. V. 40. P. 395–403.
<https://doi.org/10.18287/2412-6179-2016-40-3-395-403>
- Kunina I.A., Terekhin A.P., Gladilin S.A., Nikolaev D.P. Blind radial distortion compensation from video using fast Hough transform. *International Conference on Robotics and Machine Vision*. 2017. V. 10253. № 1025308. P. 1–7.
<https://doi.org/10.1117/12.2254867>.
- Legge G.E., Pelli D.G., Rubin G.S., Schleske M.M. Psychophysics of reading—I. Normal vision. *Vision research*. 1985. V. 25. № 2. P. 239–252.
[https://doi.org/10.1016/0042-6989\(85\)90117-8](https://doi.org/10.1016/0042-6989(85)90117-8)
- Povolotskiy M.A., Kuznetsova E.G., Khanipov T.M. Russian license plate segmentation based on dynamic time warping. *European Conference on Modelling and Simulation*. 2017. P. 285–291.
- Rodríguez-Piñero 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. V. 787410. P. 1–12.
- Sim D.G., Kwon O.K., Park R.H. Object matching algorithms using robust Hausdorff distance measures. *IEEE Transactions on image processing*. 1999. V. 8. № 3. P. 425–429.
<https://doi.org/10.1109/83.748897>
- Orrite C., Herrero J.E. Shape matching of partially occluded curves invariant under projective transformation. *Computer Vision and Image Understanding*. 2004. V. 93. № 1. P. 34–64.
<https://doi.org/10.1016/j.cviu.2003.09.005>
- Skoryukina N., Chernov T., Bulatov K., Nikolaev D.P., Arlazarov V. Snapscreen: TV-stream frame search with projectively distorted and noisy query. *Ninth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2017. V. 103410Y. P. 1–5.
<https://doi.org/10.1117/12.2268735>.
- Skoryukina N., Shemiakina J., Arlazarov V.L., Faradjev I. Document localization algorithms based on feature points and straight lines. *International Society for Optics and Photonics*. 2018. V. 106961H. P. 1–5.
<https://doi.org/10.1117/12.2311478>
- Takezawa Y., Hasegawa M., Tabbone S. Camera-captured document image perspective distortion correction using vanishing point detection based on Radon transform. *23rd International Conference on Pattern Recognition (ICPR). IEEE*. 2016. P. 3968–3974.
<https://doi.org/10.1109/ICPR.2016.7900254>.
- Wei H., Wang Y., Forman G., Zhu Y. Map matching by Fréchet distance and global weight optimization. *Technical Paper, Departement of Computer Science and Engineering*. 2013. P. 19–30.
- Xie Y., Tang G., Hoff W. Geometry-based populated chessboard recognition. *Tenth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2018. V. 1069603. P. 1–5.
- Zhang Z., He L.W. Whiteboard scanning and image enhancement. *Digital Signal Processing*. 2007. V. 17. № 2. P. 414–432.
<https://doi.org/10.1016/j.dsp.2006.05.006>
- 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. *IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 2017. №1. P. 337–342.
<https://doi.org/10.1109/ICDAR.2017.63>.

Maximal directions discrepancy as accuracy criterion of images projective normalization for optical text recognition

I. A. Konovalenko^{a,b,#}, D. V. Polevoy^{b,c,d,e}, and D. P. Nikolaev^{a,b}

^a Institute for Information Transmission Problems (Kharkevich Institute) RAS
127051 Moscow, B. Karetny per. 19, Russia

^b Smart Engines Service LLC
117312 Moscow, pr. 60-letiya Oktyabrya, 9, Russia

^c Institute for System Analysis of Federal Research Center "Computer Science and Control" RAS
117312 Moscow, pr. 60-letiya Oktyabrya, 9, Russia

^d National University of Science and Technology "MISIS"
119991 Moscow, Leninsky prospect, 4, Russia

^e Moscow Institute of Physics and Technology (State University)
141701 Dolgoprudny, Institutsky pereulok, 9, Moscow Region, Russia

#E-mail: konovalenko@iitp.ru

The application of projective normalization (a special case of orthocorrection) to photographs of documents for their further optical recognition is generally accepted. To date, a number of criteria are known for the accuracy of projective normalization. Almost all of them characterize only the coordinates discrepancy. However, the text fields of documents usually have an elongated shape, so that even with small coordinates discrepancy, large directions discrepancy are possible, which significantly affect the quality of segmentation of the field and the recognition of individual characters in it. The problem of accurate correction of directions discrepancy also arises in tomography problems if a spiral scanning scheme is used for measurement or projections are recorded in tomosynthesis schemes. To describe images projective normalization accuracy at a point, a pointwise maximum directions discrepancy is proposed. As a criterion for projective normalization accuracy of the entire image, a maximum directions discrepancy equal to the maximum pointwise maximum directions discrepancy in the region of interest is proposed. An analytical solution to the problem of calculating the pointwise maximum directions discrepancy is obtained. A hypothesis was put forward and numerically confirmed that the pointwise maximum directions discrepancy is a quasiconvex function. The theorem is proved that the supremum of a quasiconvex function on a bounded closed set is equal to the supremum on the extreme points of its convex hull. Based on the hypothesis and theorem, an analytical solution to the problem of calculating the maximum directions discrepancy on the polyhedral region of interest is proposed.

Key words: orthocorrection, perspective correction, images projective normalization, accuracy criteria, directions discrepancy, optical character recognition, mathematical programming

REFERENCES

- Balickij A.M., Savchik A.V., Gafarov R.F., Konovalenko I.A. O proektivno invariantnyh tochkah ovala s vydelennoj vneshnej prjamoj [About design-invariant points of an oval with a distinguished external line]. *Problemy pere-dachi informacii* [Information Transfer Issues]. 2017. V. 53. № 3. P. 84–89 (in Russian).
- Berezskij O.N., Berezskaja K.M. Kolichestvennaja ocenka kachestva segmentacii izobrazhenij na osnove metric [Quantitative assessment of image segmentation quality based on metrics]. *Upravljajushhie sistemy i mashiny* [Control systems and machines]. 2015. № 6. P. 59–65 (in Russian).
- Bolotova J.A., Spicyn V.G., Osina P.M. Obzor algoritmov detektirovaniya tekstovyh oblastej na izobrazhenijah i videozapisjah. [An overview of the algorithms for detecting text areas in images and videos] *Komp'juternaja optika* [Computer optics]. 2017. V. 41. № 3. P. 441–452 (in Russian).
- Buzmakov A.V., Asadchikov V.E., Zolotov D.A., Chukalina M.V., Ingacheva A.S., Krivonosov Y.S. Laboratornye rentgenovskie mikrotomografy: metody predobrabotki jeksperimental'nyh dannyh [Laboratory X-ray microtomographs: experimental data preprocessing methods]. *Izvestija RAN. Serija fizicheskaja* [Proceedings of the RAS. Physical series]. 2019. V. 83 (2). P. 194–197.
<https://doi.org/10.1134/S0367676519020066> (in Russian).
- Efimov A.I., Novikov A.I. Algoritm pojetapnogo utochenija proektivnogo preobrazovanija dlja sovmeshhenija izobrazhenij [Algorithm for step-by-step refinement of projective transformation for image matching]. *Komp'juternaja optika* [Computer optics]. 2016. V. 40. № 2. P. 258–265.
<https://doi.org/10.18287/2412-6179-2016-40-2-258-265> (in Russian).
- Zeynalov R., Velizhev A., Konushin A. Vosstanovlenie formy stranicy teksta dlja korrekcii geometricheskikh iskazhenij [Recovering the shape of a page of text for correcting geometric distortions]. *19 International Conference GraphiCon-2009*, Moscow. 2009. P. 125–128 (in Russian).
- Katamanov S.N. Avtomaticheskaja privjazka izobrazhenij geostacionarnogo sputnika MTSAT-1R [MTSAT-1R automatic geostationary satellite image linking]. *Sovremennye problemy distancionnogo zondirovaniya Zemli iz kosmosa* [Modern problems of remote sensing of the

- Earth from space]. 2007. V. 1. № 4. P. 63–68 (in Russian).
- Nikolayev P.P. Projectively invariant description of non-planar smooth figures. 1. Preliminary analysis of the problem. *Sensornye sistemy* [Sensory system]. 2016. V. 30. № 4. P. 290–311 (in Russian).
- Pritula N.E., Nikolaev P.P., Sheshkus A.V. Svravnenie dvuh algoritmov proektivno-invariantnogo raspoznavanija ploskih zamknutyh konturov s edinstvennoj vognutost'ju [Comparison of two algorithms for projectively invariant recognition of flat closed loops with a single concavity]. *Sbornik trudov ITIS-14* [Proceedings ITIS-14]. 2014. P. 367–373 (in Russian).
- Putjatin E.P., Prokopenko D.O., Pechenaja E.M. Voprosy normalizacii izobrazhenij pri proektivnyh preobrazovanijah [Image normalization issues in projective transformations]. *Radioelektronika i informatika* [Electronics and Informatics]. 1998. № 2(3). P. 82–86 (in Russian).
- Rokafellar R. *Vypuklyj analiz* [Convex analysis]. M. Mir [Peace]. 1973. V. 472 (in Russian).
- Savchik A.V., Nikolaev P.P. Metod proektivnogo sopostavlenija dlja ovalov s dvumja otmechennymi tochkami [Projective matching method for ovals with two marked points]. *Informacionnye tehnologii i vychislitel'nye sistemy* [Information Technology and Computing Systems]. 2018. № 1. C. 60–67 (in Russian).
- Holopov I.S. Algoritm korrekcii proektivnyh iskazhenij pri malovysotnoj sjomke [Projection distortion correction algorithm for low-altitude shooting]. *Komp'juternaja optika* [Computer optics]. 2017. V. 41. № 2. S 284–290 (in Russian).
- Shapiro L., Stokman D., Boguslavskij A.A., Sokolov S.M. Komp'juternoe zrenie [Computer vision]. M. BINOM. *Laboratorija znanij* [BINOMIAL. Knowledge laboratory]. 2013. 763 p. (in Russian).
- Shemiakina J.A., Zhukovsky A.E., Faradjev I.A. The research of the algorithms of a projective transformation calculation in the problem of planar object targeting by feature points. *Iskusstvennyj intellekt i prinjatie reshenij* [Artificial Intelligence and Decision Making]. 2017. № 1. P. 43–49 (in Russian).
- Judin D.B. Matematicheskie metody upravlenija v uslovijah nepolnoj informacii [Mathematical control methods in conditions of incomplete information]. *Izdatel'skaya grupa URSS* [URSS Publishing Group]. Moscow, Russia. 2010. 400 p. (in Russian).
- 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. V. 11. № 2. P. 14–28. <https://doi.org/10.14529/mmp180202>.
- Arvind C.S., Mishra R., Vishal K., Gundimeda V. Vision based speed breaker detection for autonomous vehicle. Tenth International Conference on Machine Vision (ICMV): *International Society for Optics and Photonics*. 2018. V. 106960E. P. 1–9. <https://doi.org/10.1117/12.2311315>.
- Awal A.M., Ghanmi N., Sicre R., Furon T. Complex document classification and localization application on identity document images. *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. – IEEE. 2017. V. 1. P. 426–431. <https://doi.org/10.1109/ICDAR.2017.77>.
- Bezmaternykh P.V., Nikolaev D.P., Arlazarov V.L. Textual blocks rectification method based on fast Hough transform analysis in identity documents recognition. *Tenth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2018. V. 1069606. P. 1–6. <https://doi.org/10.1117/12.2310162>.
- Calore E., Pedersini F., Frosio I. Accelerometer based horizon and keystone perspective correction. *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. 2012. P. 205–209. <https://doi.org/10.1109/I2MTC.2012.6229434>.
- 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. P. 339–346.
- Dubuisson M.P., Jain A.K. A modified Hausdorff distance for object matching. *Proceedings of 12th international conference on pattern recognition. IEEE*. 1994. V. 1. P. 566–568. <https://doi.org/10.1109/ICPR.1994.576361>.
- Fréchet M.M. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo* (18841940). 1906. V. 22. № 1. P. 1–72.
- Hsu S.C., Sawhney H.S. Influence of global constraints and lens distortion on pose and appearance recovery from a purely rotating camera. *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98* (Cat. No. 98EX201). 1998. P. 154–159. <https://doi.org/10.1109/ACV.1998.732873>.
- Huttenlocher D.P., Klanderman G.A., Rucklidge W.J. Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*. 1993. V. 15. № 9. P. 850–863. <https://doi.org/10.1109/34.232073>.
- Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*. 1901. V. 37. P. 547–579.
- 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*. 2001. P. 90–95.
- Karpenko S., Konovalenko I., Miller A., Miller B., Nikolaev D. UAV control on the basis of 3D landmark bearing-only observations. *Sensors*. 2015. V. 15. № 12. P. 29802–29820. <https://doi.org/10.3390/s151229768>.
- Konovalenko I.A., Shemiakina J.A. Error values analysis for inaccurate projective transformation of a quadrangle. *Journal of Physics: Conference Series*. – IOP Publishing. 2018. V. 1096. № 1. P. 1–9. <https://doi.org/10.1088/1742-6596/1096/1/012038>.
- Kunina I.A., Gladilin S.A., Nikolaev D.P. Blind radial distortion compensation in a single image using fast Hough transform. *Computer optics*. 2016. V. 40. P. 395–403. <https://doi.org/10.18287/2412-6179-2016-40-3-395-403>.
- Kunina I.A., Terekhin, A.P., Gladilin, S.A., Nikolaev D.P. Blind radial distortion compensation from video using

- fast Hough transform. *International Conference on Robotics and Machine Vision*. 2017. V. 10253. № 1025308. P. 1–7.
<https://doi.org/10.1117/12.2254867>.
- Legge G.E., Pelli D.G., Rubin G.S., Schleske M.M. Psychophysics of reading—I. *Normal vision. Vision research*. 1985. V. 25. № 2. P. 239–252.
[https://doi.org/10.1016/0042-6989\(85\)90117-8](https://doi.org/10.1016/0042-6989(85)90117-8).
- Povolotskiy M.A., Kuznetsova E.G., Khanipov T.M. Russian license plate segmentation based on dynamic time warping. *European Conference on Modelling and Simulation*. 2017. P. 285–291.
- Rodríguez-Piñero 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. V. 787410. P. 1–12.
- Sim D.G., Kwon O.K., Park R.H. Object matching algorithms using robust Hausdorff distance measures. *IEEE Transactions on image processing*. 1999. V. 8. № 3. P. 425–429.
<https://doi.org/10.1109/83.748897>.
- Orrite C., Herrero J.E. Shape matching of partially occluded curves invariant under projective transformation. *Computer Vision and Image Understanding*. 2004. V. 93. № 1. P. 34–64.
<https://doi.org/10.1016/j.cviu.2003.09.005>.
- Skoryukina N., Chernov T., Bulatov K., Nikolaev D.P., Arlazarov V. Screenshot: TV-stream frame search with projectively distorted and noisy query. *Ninth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2017. V. 103410Y. P. 1–5.
<https://doi.org/10.1117/12.2268735>.
- Skoryukina N., Shemiakina J., Arlazarov V.L., Faradjev I. Document localization algorithms based on feature points and straight lines. *International Society for Optics and Photonics*. 2018. V. 106961H. P. 1–5.
<https://doi.org/10.1117/12.2311478>.
- Takezawa Y., Hasegawa M., Tabbone S. Camera-captured document image perspective distortion correction using vanishing point detection based on Radon transform. 2016 *23rd International Conference on Pattern Recognition (ICPR)*. *IEEE*. 2016. P. 3968–3974.
<https://doi.org/10.1109/ICPR.2016.7900254>.
- Wei H., Wang Y., Forman G., Zhu Y. Map matching by Fréchet distance and global weight optimization. *Technical Paper, Departement of Computer Science and Engineering*. 2013. P. 19–30.
- Xie Y., Tang G., Hoff W. Geometry-based populated chessboard recognition. *Tenth International Conference on Machine Vision (ICMV): International Society for Optics and Photonics*. 2018. V. 1069603. P. 1–5.
- Zhang Z., He L.W. Whiteboard scanning and image enhancement. *Digital Signal Processing*. 2007. V. 17. № 2. P. 414–432.
<https://doi.org/10.1016/j.dsp.2006.05.006>.
- 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. *IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 2017. № 1. P. 337–342.
<https://doi.org/10.1109/ICDAR.2017.63>.