Optimal affine image normalization approach for optical character recognition

I.A. Konovalenko [1,2], V.V. Kokhan [1,2], D.P. Nikolaev [1,2]
[1] Institute for Information Transmission Problems RAS, 127051, Moscow, Russia, Bolshoy Karetny per. 19, bld. 1,
[2] Smart Engines, 117312, Moscow, Russia, pr-t 60-letiya Oktyabrya, 9

Abstract

Optical character recognition (OCR) in images captured from arbitrary angles requires preliminary normalization, i.e. a geometric transformation resulting in an image as if it was captured at an angle suitable for OCR. In most cases, a surface containing characters can be considered flat, and a pinhole model can be adopted for a camera. Thus, in theory, the normalization should be projective. Usually, the camera optical axis is approximately perpendicular to the document surface, so the projective normalization can be replaced with an affine one without a significant loss of accuracy. An affine image transformation is performed significantly faster than a projective normalization, which is important for OCR on mobile devices. In this work, we propose a fast approach for image normalization. It utilizes an affine normalization instead of a projective one if there is no significant loss of accuracy. The approach is based on a proposed criterion for the normalization accuracy: root mean square (RMS) coordinate discrepancies over the region of interest (ROI). The problem of optimal affine normalization according to this criterion is considered. We have established that this unconstrained optimization is quadratic and can be reduced to a problem of fractional quadratic functions integration over the ROI. The latter was solved analytically in the case of OCR where the ROI consists of rectangles. The proposed approach is generalized for various cases when instead of the affine transform its special cases are used: scaling, translation, shearing, and their superposition, allowing the image normalization procedure to be further accelerated.

Keywords: optical character recognition, image registration, image normalization, coordinate discrepancy, projective transformation, affine transformation, approximation, optimization, symbolic computation.

Citation: Konovalenko IA, Kokhan VV, Nikolaev DP. Optimal affine image normalization approach for optical character recognition. Computer Optics 2021; 45(1): 90-100. DOI: 10.18287/2412-6179-CO-759.

Acknowledgments: This work was partially financially supported by the Russian Foundation for Basic Research, projects 18-29-26035 and 17-29-03370.

Introduction

Projective image normalization

Optical character recognition (OCR) in images captured from arbitrary angles requires preliminary normalization, i.e. geometric transformation resulting in an image as if it was captured from the angle suitable for OCR. In most cases, a surface containing characters can be considered flat, and a pinhole model can be adopted for a camera. Thus, in theory, the normalization should be projective. The latter is commonly employed as a part of image preprocessing for various computer vision tasks, such as document OCR [1, 2, 3, 4, 5], vehicle license plate recognition [6], TV-stream recognition based on a picture of a TV screen [7], chessboard recognition [8], artificial on-road obstacles detection [9], object detection using shape features (detection of the shape of an object within an image and matching that shape with an object from database) [10, 11, 12, 13, 14, 15], surface parameters monitored from satellites (time-temporal variability of sea surface temperature, determining the velocity of the cloud masses motion, etc.) [16], reconstruction of plans and maps from the aerial photographs [17, 18], and many more. In addition, the projective normalization of photographs of documents helps human perception [19].

Affine approximation of projective normalization

Usually, the camera optical axis is approximately perpendicular to the document surface. In such cases, a projection model of the affine camera can be utilized [20], and a projective normalization can be replaced with a commonly used affine normalization without significant loss of accuracy [21, 22]. The affine image transformation is performed significantly faster than the projective normalization [22, 23], which is helpful for fast image normalization. The latter is important for the OCR on mobile devices [24].

The idea of the replacement of the projective transformation with the affine one in practice was mentioned in [25] back in 1985. This property was implemented in [26] for the simplification of the mathematical calculations. The affine approximation is commonly used in image completion [27] and rendering [23, 28, 29]. In [30] the projective transformation is replaced with the simpler affine transformation in order to avoid overfitting. A similar idea is utilized in «weak-perspective projection» [31, 32, 33], where the approximation is partial. Use of the affine invariant methods instead of significantly more complicated projective invariant methods is common in keypoints technology [34, 35, 36], as well as in the related

problem of salient region detection [37], and both of these methods are essentially camera angle invariant. There is also a division into affine and projective methods in stereo reconstruction [38]. The utilization of affine transformation for image rendering and normalization results in loss of accuracy [22, 39], but the accuracy was not formally introduced.

Affine approximation of the given projective normalization aiming to accelerate the latter is considered for the first time in this work.

Definitions and notation

Let I_{input} be an input image (usually a photograph) for the normalization. Let its known projective normalization be a perfect normalization H. Let an image formed as the result of the application of H to I_{input} be a projectively normalized image I_{proj} (see Fig. 1). An arbitrary affine approximation of the projective normalization H is denoted as A: $\mathrm{A} = \hat{\mathrm{H}}$. Thus A is the affine normalization of the image I_{input}. The resulting image I_{affin} of A applied to I_{input} is an affinely normalized image.

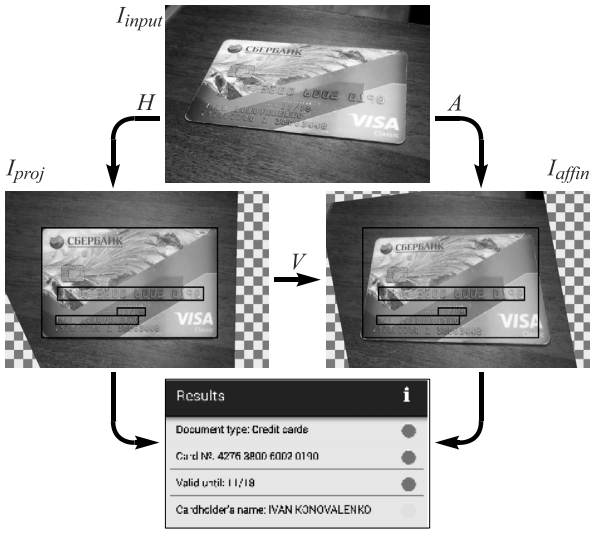

Fig. 1. The general scheme of transformations, where I_{input} is an image of the document captured from an arbitrary angle, I_{proj} is a projectively normalized image, I_{affin} is an affinely normalized image, and the result of the OCR

Let $\mathbf{r} \overset{\text{def}}{=} [x\ y]^T$ be Cartesian coordinates of pixels on the plane of I_{proj}. We define the residual projective distortion as

$$\mathrm{V} \overset{\text{def}}{=} \mathrm{A}\mathrm{H}^{-1}, \tag{1}$$

which for each point of the scene transforms coordinates $\mathbf{r}$ of its image on I_{proj} into coordinates $\mathrm{V}(\mathbf{r})$ of its image on I_{affin}.

Ideally, the residual distortion V is an identical transformation. For the formalization of pointwise error of affine normalization we define the coordinate discrepancies [40] (see Fig. 2) as

$$\mathrm{d}(\mathbf{r}) \overset{\text{def}}{=} \| \mathbf{r} - \mathrm{V}(\mathbf{r}) \|_2 . \tag{2}$$

In some cases, it is possible to evaluate beforehand which part of the projectively normalized image I_{proj} is of interest. Such region of interest (ROI) is denoted as $\mathrm{R} \subset \mathbb{R}^2$. Otherwise, R denotes the entire I_{proj}.

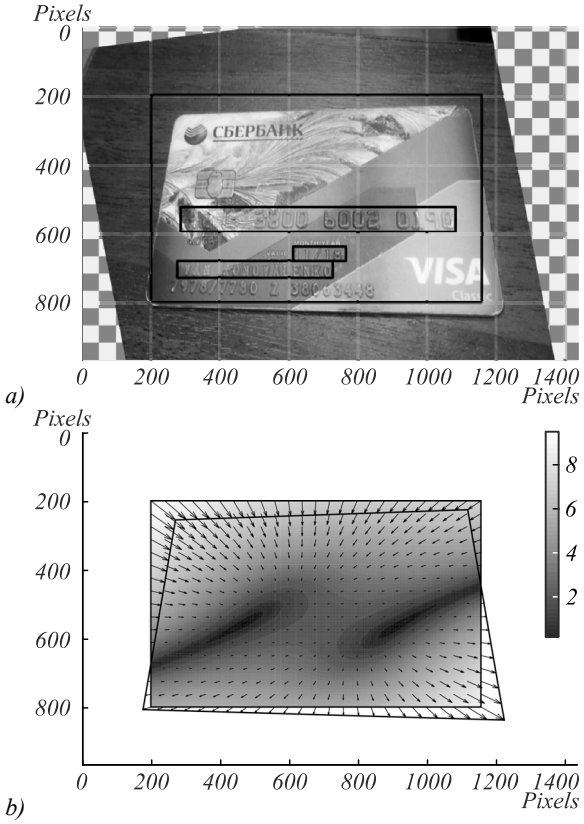

Fig. 2. The coordinate discrepancies. a) the affinely normalized image I_{affin}; black frames indicate the ideal positions of the text fields; b) a shift vector field $\mathrm{V}(\mathbf{r})-\mathbf{r}$, $\mathbf{r} \in R$; the shades of grey illustrate the square root of coordinate discrepancies $\sqrt{\mathrm{d}(\mathbf{r})}$

1. Root mean square criterion of normalization accuracy

As a criterion of normalization accuracy, we choose the widely used criterion of root mean square (RMS) coordinate discrepancies. In cases of ROI with finite non-zero area $0 < S(R) < \infty$ and non-empty finite ROI $0 < |R| < \infty$ the criterion is defined as follows:

$$L_2(\mathrm{d}; R) \overset{\text{def}}{=} \begin{cases} \sqrt{\dfrac{1}{S(R)} \displaystyle\int_R \mathrm{d}^2(\mathbf{r})\ \mathrm{d}\mathbf{r}} & \text{for } 0 < S(R) < \infty, \\ \sqrt{\dfrac{1}{|R|} \displaystyle\sum_{\mathbf{r} \in R} \mathrm{d}^2(\mathbf{r})} & \text{for } 0 < |R| < \infty. \end{cases} \tag{3}$$

Such criterion was used, for example, for the automatic normalization of distortion caused by lens distortion and camera movement [41]. The same criterion was also employed for the calculation of the accuracy of the aligned image formation via projectors matrix [42]. Using definitions (1) and (2) we establish the dependence of criterion on the affine transformation A:

$$L_2(\mathrm{A}, \mathrm{H}; R) \overset{\text{def}}{=} L_2(\mathrm{d}; R), \quad \mathrm{d}(\mathbf{r}) = \| \mathbf{r} - \mathrm{A}\mathrm{H}^{-1}(\mathbf{r}) \|_2 . \tag{4}$$

2. Optimal affine image normalization

2.1. Problem formulation

Now, as the criterion of normalization accuracy is set, we can formulate a problem of search for the optimal affine approximation of the projective normalization H:

$$\mathrm{A}^* \overset{\text{def}}{=} \arg\min_{\mathrm{A}} L_2(\mathrm{A},\mathrm{H};R). \tag{5}$$

We will also refer to it as the optimal affine normalization. The correspondent optimum is denoted as

$$L_2^* \overset{\text{def}}{=} \min_{\mathrm{A}} L_2(\mathrm{A},\mathrm{H};R) = L_2(\mathrm{A}^*,\mathrm{H};R). \tag{6}$$

The projective normalization H is parametrized by the homography matrix $H \overset{\text{def}}{=} (h_{ij}) \in \mathbb{R}^{3\times 3}$:

$$\mathbf{r} = \mathrm{H}(\mathbf{r}') \overset{\text{def}}{=} \frac{\begin{bmatrix} h_{11}x' + h_{12}y' + h_{13} \\ h_{21}x' + h_{22}y' + h_{23} \end{bmatrix}}{h_{31}x' + h_{32}y' + h_{33}},$$

where $\mathbf{r}' \overset{\text{def}}{=} [x'\ y']^T$ are Cartesian coordinates of pixels on I_{input} image surface. Let an inverse transformation for H transformation be $\mathrm{P} \overset{\text{def}}{=} \mathrm{H}^{-1}$ and we parametrize it by matrix $P \overset{\text{def}}{=} (p_{ij}) \in \mathbb{R}^{3\times 3}$:

$$\mathbf{r}' = \mathrm{P}(\mathbf{r}) \overset{\text{def}}{=} \frac{\begin{bmatrix} p_{11}x + p_{12}y + p_{13} \\ p_{21}x + p_{22}y + p_{23} \end{bmatrix}}{p_{31}x + p_{32}y + p_{33}}, \tag{7}$$

then $P \sim H^{-1}$. Because matrices P and H are homogenous we assume

$$P = H^{-1}. \tag{8}$$

The affine transformation A is parametrized by matrix $A \overset{\text{def}}{=} (a_{ij}) \in \mathbb{R}^{2\times 3}$:

$$\mathrm{A}(\mathbf{r}) \overset{\text{def}}{=} A\,[x\ y\ 1]^T. \tag{9}$$

Thus, problem (5) of the optimal transformation search can be formulated as the problem of optimal matrix search

$$A^* = \arg\min_{A} \begin{cases} \int\limits_R \left\| \mathbf{r} - A\begin{bmatrix} \mathrm{P}(\mathbf{r}) \\ 1 \end{bmatrix} \right\|_2^2 \mathrm{d}\mathbf{r} & \text{for } 0<S(R)<\infty, \\ \sum\limits_{\mathbf{r}\in R} \left\| \mathbf{r} - A\begin{bmatrix} \mathrm{P}(\mathbf{r}) \\ 1 \end{bmatrix} \right\|_2^2 & \text{for } 0<|R|<\infty. \end{cases} \tag{10}$$

Earlier in [43] we proved that this problem is convex.

2.2. The applicability limits

Consider function

$$Z(\mathbf{r}) \overset{\text{def}}{=} p_{31}x + p_{32}y + p_{33}. \tag{11}$$

The line $Z(\mathbf{r})=0$ on I_{proj} image surface is denoted as the horizon. Let us consider ROI R which does not lie strictly on one part outlined by the horizon. Points on the horizon turn the denominator of the transformation (7) into zero, which corresponds to them being infinitely remote on the input image I_{input} plane. Hence these points cannot be present in I_{input} image because of its finite size. In reality, these points of a scene are situated in the $\pi/2$ angle of camera view. Points that belong to the different sides of the horizon cannot be simultaneously present in I_{input} image, because points that belong to one of these sides are situated in $>\pi/2$ angle of camera view, i.e. located behind the camera. Thus, at least a part of the ROI is absent in the input I_{input} image. In this case, the RMS criterion of accuracy (4) is meaningless. Hence we will consider only cases when the ROI lies strictly on one of the sides outlined by the horizon:

$$\begin{bmatrix} Z(\mathbf{r}\in R) < 0, \\ Z(\mathbf{r}\in R) > 0. \end{bmatrix} \tag{12}$$

This condition also guarantees the correctness of the RMS accuracy criterion definition (4).

2.3. ROI of non-zero finite area

Let us consider the ROI with the non-zero finite area, then from (10) follows

$$A^* = \arg\min_{A} \int_R \left\| \mathbf{r} - A\begin{bmatrix} \mathrm{P}(\mathbf{r}) \\ 1 \end{bmatrix} \right\|_2^2 \mathrm{d}\mathbf{r}. \tag{13}$$

We will express the affine transformation matrix A as the vector $\mathbf{a} \overset{\text{def}}{=} (a_i) \in \mathbb{R}^6$:

$$\mathbf{a}(A) \overset{\text{def}}{=} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} \Rightarrow A(\mathbf{a}) = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{bmatrix}. \tag{14}$$

Let us specify the transformation P through its components:

$$\mathrm{P} \overset{\text{def}}{=} \begin{bmatrix} \mathrm{P}_x & \mathrm{P}_y \end{bmatrix}^T, \tag{15}$$

and introduce a matrix function Q:

$$\mathrm{Q} \overset{\text{def}}{=} \begin{bmatrix} \mathrm{P}_x & \mathrm{P}_y & 1 & & & \\ & & & \mathrm{P}_x & \mathrm{P}_y & 1 \end{bmatrix}. \tag{16}$$

Then

$$A\begin{bmatrix} \mathrm{P} \\ 1 \end{bmatrix} = \mathrm{Q}\mathbf{a},$$

which allows the problem vectorization (13) to be defined as follows:

$$A^* = A(\mathbf{a}^*), \quad \mathbf{a}^* \overset{\text{def}}{=} \arg\min_{\mathbf{a}} \int_R \| \mathbf{r} - \mathrm{Q}(\mathbf{r})\mathbf{a} \|_2^2\, \mathrm{d}\mathbf{r}. \tag{17}$$

Note that the target function of the problem (17) is quadratic:

$$\int_R \| \mathbf{r} - \mathrm{Q}(\mathbf{r})\mathbf{a} \|_2^2 \, \mathrm{d}\mathbf{r} = K^{\{0\}} - 2K^{\{1\}}\mathbf{a} + \mathbf{a}^T K^{\{2\}}\mathbf{a}, \qquad (18)$$

$$\text{where } K^{\{k\}} = \int_R f^{\{k\}}(\mathbf{r})\,\mathrm{d}\mathbf{r}, \qquad (19)$$

$$\text{where } \begin{aligned} f^{\{0\}}(\mathbf{r}) &\overset{\text{def}}{=} \mathbf{r}^T\mathbf{r}, \\ f^{\{1\}}(\mathbf{r}) &\overset{\text{def}}{=} \mathbf{r}^T\mathrm{Q}(\mathbf{r}), \\ f^{\{2\}}(\mathbf{r}) &\overset{\text{def}}{=} \mathrm{Q}^T(\mathbf{r})\mathrm{Q}(\mathbf{r}). \end{aligned} \qquad (20)$$

We will refer to the coefficients K as the target coefficients. As was shown above, the target coefficients are defined by the homography matrix and the ROI:

$$K = \mathrm{K}(H, R). \qquad (21)$$

If the target coefficients are calculated, the problem (17) can be presented as

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \left(K^{\{0\}} - 2K^{\{1\}}\mathbf{a} + \mathbf{a}^T K^{\{2\}}\mathbf{a} \right),$$

and can be solved analytically:

$$\mathbf{a}^* = \left(K^{\{2\}} \right)^{-1} \left(K^{\{1\}} \right)^T. \qquad (22)$$

Thus, the problem of the unconstrained normalization (13) is quadratic and can be reduced to the problem of the fractional quadratic functions integration over the ROI. Obviously, for an arbitrary ROI this integration can be performed only numerically.

2.4. Non-empty finite ROI

Similar reasoning can be suggested for the non-empty finite ROI R. In this case, according to (10):

$$K^{\{k\}} = \sum_{\mathbf{r}\in R} f^{\{k\}}(\mathbf{r}), \qquad (23)$$

while the definition (20) of functions f and the expression (22) for analytical calculation of $\mathbf{a}^*$ are preserved. Hence the RMS criterion (3) can be calculated as

$$L_2(\mathrm{A}, \mathrm{H}; R) = D^{-1/2} \left(K^{\{0\}} - 2K^{\{1\}}\mathbf{a} + \mathbf{a}^T K^{\{2\}}\mathbf{a} \right)^{1/2}, \qquad (24)$$

$$\text{where } D \overset{\text{def}}{=} \begin{cases} S(R) & \text{for } 0 < S(R) < \infty, \\ |R| & \text{for } 0 < |R| < \infty. \end{cases}$$

Thus, in all considered cases (3) the optimal affine normalization is calculated according to the general Algorithm 1.

Notes on Step 1 regarding the calculation of the target coefficients $K = \mathrm{K}(H, R)$. The cases of the non-empty finite ROI and the ROI of non-zero finite areas are discussed above. Let us specify the corresponding Algorithms (2 and 3) for the target coefficients calculation:

Algorithm 1. *Algorithm of the optimal affine image normalization search*

Input:
- matrix $H \in \mathbb{R}^{3\times3}$ of projective normalization H,
- ROI $\mathrm{R} \subset \mathbb{R}^2$: $0 < S(R) < \infty$ or $0 < |R| < \infty$.

Output:
- matrix $A^* \in \mathbb{R}^{2\times3}$ of optimal affine approximation H on R: (9),
- the optimal value of RMS accuracy criterion L_2^*: (6).

Step 1. Based on H and R target coefficients are calculated $K = \mathrm{K}(H, R)$.
Step 2. $\mathbf{a}^*$ is calculated: (22).
Step 3. $A^* = A(\mathbf{a}^*)$ is calculated: (14).
Step 4. $L_2^* = L_2(\mathrm{A}^*, \mathrm{H}; R)$ is calculated: (24).

Algorithm 2. *Calculation of the target coefficients for the non-empty finite ROI*

Input:
- matrix $H \in \mathbb{R}^{3\times3}$ of projective normalization H,
- non-empty finite ROI $0 < |R| < \infty$.

Output: Target coefficients $K = \mathrm{K}(H, R)$.
Step 1. Matrix $P = (p_{ij})$ is calculated: (8).
Step 2. P_x and P_y are defined: (7), (15).
Step 3. Function Q is defined: (16).
Step 4. Functions f are defined: (20).
Step 5. Target coefficients K are calculated: (23).

Algorithm 3. *Numerical estimation of the target coefficients for the ROI of non-zero finite area*

Input:
- matrix $H \in \mathbb{R}^{3\times3}$ of projective normalization H,
- ROI $\mathrm{R} \subset \mathbb{R}^2$ of the non-zero finite area: $0 < S(R) < \infty$.

Output: Numerical estimation of the target coefficients $K = \mathrm{K}(P, R)$.
Step 1. Matrix $P = (p_{ij})$ is calculated: (8).
Step 2. P_x и P_y are defined: (7), (15).
Step 3. Function Q is defined: (16).
Step 4. Functions f are defined: (20).
Step 5. Set $\{\mathbf{r}_i\}_{i=1}^n$ of uniformly distributed on R points is generated.
Step 6. Assignment $R := \{\mathbf{r}_i\}_{i=1}^n$.
Step 7. Target coefficients K are computationally evaluated: (23).

In order to get the conventional statistical estimation, result of (23) should be multiplied by $S(R)/n$ at the final step of Algorithm 3. This multiplication is skipped intentionally, because on the one hand, it does not change the output of the Algorithm 1 (see expression (22)), and on the other hand, the accurate calculation of the area $S(R)$ in special cases complicates the Algorithm 3, and generally might not be even possible.

Further we will analytically calculate the target coefficients K for some special cases of the ROI R with non-zero finite area.

2.5. Orthotropic rectangular ROI

In computer vision applications there is a particularly important case of the orthotropic rectangular ROI R:

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

Let us introduce the antiderivatives

$$F^{\{k\}}(\mathbf{r})\overset{\text{def}}{=}\int f^{\{k\}}(\mathbf{r})\,\mathrm{d}\mathbf{r},$$

then by the Newton–Leibniz axiom, expressions (19) can be written as

$$K^{\{k\}}=F^{\{k\}}(x,y)\big|_{x_1}^{x_2}\big|_{y_1}^{y_2}, \tag{25}$$

where $F(x,y)\big|_{x_1}^{x_2}\big|_{y_1}^{y_2}\overset{\text{def}}{=}F(x_1,y_1)+F(x_2,y_2)-$
$-F(x_1,y_2)-F(x_2,y_1).$

Now in order to calculate (19) we have to find the antiderivatives F. Let us introduce the changes of variables:

$$\begin{aligned}
c_1&=2p_{11}p_{32}^2-2p_{12}p_{31}p_{32},\\
c_2&=2p_{12}p_{31}^2-2p_{11}p_{31}p_{32},\\
c_3&=2p_{11}p_{32}p_{33}-2p_{13}p_{31}p_{32}+\\
&\quad+2p_{11}p_{32}p_{33}-2p_{12}p_{31}p_{33},\\
c_4&=p_{11}p_{31}^2,\\
c_5&=2p_{13}p_{31}^2-2p_{11}p_{31}p_{33},\\
c_6&=2p_{11}p_{33}^2-2p_{13}p_{31}p_{33},\\
c_7&=2p_{21}p_{32}^2-2p_{22}p_{31}p_{32},\\
c_8&=2p_{22}p_{31}^2-2p_{21}p_{31}p_{32},\\
c_9&=2p_{21}p_{32}p_{33}-2p_{23}p_{31}p_{32}+\\
&\quad+2p_{21}p_{32}p_{33}-2p_{22}p_{31}p_{33},\\
c_{10}&=p_{21}p_{31}^2,\\
c_{11}&=2p_{23}p_{31}^2-2p_{21}p_{31}p_{33},\\
c_{12}&=2p_{21}p_{33}^2-2p_{23}p_{31}p_{33},
\end{aligned} \tag{26}$$

$$\begin{aligned}
c_{13}&=2p_{12}p_{31}^2-2p_{11}p_{31}p_{32},\\
c_{14}&=2p_{11}p_{32}^2-2p_{12}p_{31}p_{32},\\
c_{15}&=2p_{12}p_{31}p_{33}-2p_{13}p_{32}p_{31}+\\
&\quad+2p_{12}p_{31}p_{33}-2p_{11}p_{32}p_{33},\\
c_{16}&=p_{12}p_{32}^2,\\
c_{17}&=2p_{13}p_{32}^2-2p_{12}p_{32}p_{33},\\
c_{18}&=2p_{12}p_{33}^2-2p_{13}p_{32}p_{33},\\
c_{19}&=2p_{22}p_{31}^2-2p_{21}p_{31}p_{32},\\
c_{20}&=2p_{21}p_{32}^2-2p_{22}p_{31}p_{32},\\
c_{21}&=2p_{22}p_{31}p_{33}-2p_{23}p_{32}p_{31}+\\
&\quad+2p_{22}p_{31}p_{33}-2p_{21}p_{32}p_{33},\\
c_{22}&=p_{22}p_{32}^2,\\
c_{23}&=2p_{23}p_{32}^2-2p_{22}p_{32}p_{33},\\
c_{24}&=2p_{22}p_{33}^2-2p_{23}p_{32}p_{33},
\end{aligned} \tag{27}$$

$$\begin{aligned}
c_{25}&=p_{12}p_{31}-p_{11}p_{32},\\
c_{26}&=p_{13}p_{31}-p_{11}p_{33},\\
c_{27}&=p_{11}p_{31},\\
c_{28}&=p_{22}p_{31}-p_{21}p_{32},\\
c_{29}&=p_{23}p_{31}-p_{21}p_{33},\\
c_{30}&=p_{21}p_{31},\\
c_{31}&=p_{11}^2p_{31},\\
c_{32}&=p_{12}^2p_{31}^2+p_{11}^2p_{32}^2-2p_{11}p_{12}p_{31}p_{32},\\
c_{33}&=2p_{12}p_{13}p_{31}^2+2p_{11}^2p_{32}p_{33}-\\
&\quad-2p_{11}p_{12}p_{31}p_{33}-2p_{11}p_{13}p_{31}p_{32},\\
c_{34}&=p_{13}^2p_{31}^2+p_{11}^2p_{33}^2-2p_{11}p_{13}p_{31}p_{33},\\
c_{35}&=2p_{11}p_{12}p_{31}-2p_{11}^2p_{32},\\
c_{36}&=2p_{11}p_{13}p_{31}-2p_{11}^2p_{33},\\
c_{37}&=p_{21}^2p_{31},\\
c_{38}&=p_{22}^2p_{31}^2+p_{21}^2p_{32}^2-\\
&\quad-2p_{21}p_{22}p_{31}p_{32},
\end{aligned} \tag{28}$$

$$\begin{aligned}
c_{39}&=2p_{22}p_{23}p_{31}^2+2p_{21}^2p_{32}p_{33}-\\
&\quad-2p_{21}p_{22}p_{31}p_{33}-2p_{21}p_{23}p_{31}p_{32},\\
c_{40}&=p_{23}^2p_{31}^2+p_{21}^2p_{33}^2-2p_{21}p_{23}p_{31}p_{33},\\
c_{41}&=2p_{21}p_{22}p_{31}-2p_{21}^2p_{32},\\
c_{42}&=2p_{21}p_{23}p_{31}-2p_{21}^2p_{33},
\end{aligned} \tag{29}$$

$$\begin{aligned}
c_{43}&=p_{12}p_{22}p_{31}^2-p_{12}p_{21}p_{31}p_{32}-\\
&\quad-p_{11}p_{22}p_{31}p_{32}+p_{11}p_{21}p_{32}^2,\\
c_{44}&=p_{12}p_{23}p_{31}^2-p_{12}p_{21}p_{31}p_{33}+\\
&\quad+p_{13}p_{22}p_{31}^2-p_{13}p_{21}p_{31}p_{32}-\\
&\quad-p_{11}p_{23}p_{31}p_{32}+p_{11}p_{21}p_{32}p_{33}-\\
&\quad-p_{11}p_{22}p_{31}p_{33}+p_{11}p_{21}p_{32}p_{33},\\
c_{45}&=p_{13}p_{23}p_{31}^2-p_{13}p_{21}p_{31}p_{33}-\\
&\quad-p_{11}p_{23}p_{31}p_{33}+p_{11}p_{21}p_{33}^2,\\
c_{46}&=p_{11}p_{22}p_{31}-2p_{11}p_{21}p_{32}+\\
&\quad+p_{12}p_{21}p_{31},\\
c_{47}&=p_{11}p_{23}p_{31}-2p_{11}p_{21}p_{33}+\\
&\quad+p_{13}p_{21}p_{31},\\
c_{48}&=p_{11}p_{21}p_{31},
\end{aligned} \tag{30}$$

$$\begin{aligned}
z_x(x)&=p_{31}x+p_{33},\\
z_y(y)&=p_{32}y+p_{33},\\
l(\mathbf{r})&=\log Z(\mathbf{r}).
\end{aligned} \tag{31}$$

The antiderivative of $\mathbf{r}^T\mathbf{r}$ is

$$F^{\{0\}}(\mathbf{r})=\int\left(\mathbf{r}^T\mathbf{r}\right)\mathrm{d}\mathbf{r}=\frac{x(x^2+y^2)y}{3}. \tag{32}$$

The antiderivatives of $\mathbf{r}^T\mathrm{Q}(\mathbf{r})$ are

$$\begin{aligned}
F^{\{1\}}(\mathbf{r})&=\int\mathbf{r}^T\mathrm{Q}(\mathbf{r})\,\mathrm{d}\mathbf{r}=\\
&=\iint\left[x\mathrm{P}_x(\mathbf{r})\;\;x\mathrm{P}_y(\mathbf{r})\;\;x\;\;y\mathrm{P}_x(\mathbf{r})\;\;y\mathrm{P}_y(\mathbf{r})\;\;y\right]\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

$$
\begin{aligned}
F_1^{\{1\}}(\mathbf{r}) &= \frac{1}{2p_{31}^3}\Big(\frac{1}{18p_{32}^3}c_1(6(p_{32}^3y^3+z_x(x)^3)l(\mathbf{r})+p_{32}y(-2p_{32}^2y^2+3p_{32}yz_x(x)-6z_x(x)^2))+y^2x\frac{c_2}{2}+\\
&+\frac{c_3}{4p_{32}^2}(p_{32}y(2z_x(x)-p_{32}y)-2(z_x(x)^2-p_{32}^2y^2)l(\mathbf{r})+yx^2c_4+yxc_5+c_6\,p_{32}y+z_x(x))\frac{l(\mathbf{r})}{p_{32}}-c_6y),\\
F_2^{\{1\}}(\mathbf{r}) &= \frac{1}{2p_{31}^3}\Big(\frac{1}{18p_{32}^3}c_7(6(p_{32}^3y^3+z_x(x)^3)l(\mathbf{r})+p_{32}y(-2p_{32}^2y^2+3p_{32}yz_x(x)-6z_x(x)^2))+y^2x\frac{c_8}{2}+\\
&+\frac{c_9}{4p_{32}^2}(p_{32}y(2z_x(x)-p_{32}y)-2(z_x(x)^2-p_{32}^2y^2)l(\mathbf{r}))+yx^2c_{10}+yxc_{11}+c_{12}(p_{32}y+z_x(x))\frac{l(\mathbf{r})}{p_{32}}-c_{12}y),\\
F_3^{\{1\}}(\mathbf{r}) &= \frac{x^2y}{2},
\end{aligned}
\tag{33}
$$

$$
\begin{aligned}
F_4^{\{1\}}(\mathbf{r}) &= \frac{1}{2p_{32}^3}\Big(\frac{1}{18p_{31}^3}c_{13}(6(p_{31}^3x^3+z_y(y)^3)l(\mathbf{r})+p_{31}x(-2p_{31}^2x^2+3p_{31}xz_y(y)-6z_y(y)^2))+x^2y\frac{c_{14}}{2}+\\
&+\frac{c_{15}}{4p_{31}^2}(p_{31}x(2z_y(y)-p_{31}x)-2(z_y(y)^2-p_{31}^2x^2)l(\mathbf{r}))+xy^2c_{16}+yxc_{17}+c_{18}(p_{31}x+z_y(y))\frac{l(\mathbf{r})}{p_{31}}-c_{18}x),\\
F_5^{\{1\}}(\mathbf{r}) &= \frac{1}{2p_{32}^3}\Big(\frac{1}{18p_{31}^3}c_{19}(6(p_{31}^3x^3+z_y(y)^3)l(\mathbf{r})+p_{31}x(-2p_{31}^2x^2+3p_{31}xz_y(y)-6z_y(y)^2))+x^2y\frac{c_{20}}{2}+\\
&+\frac{c_{21}}{4p_{31}^2}(p_{31}x(2z_y(y)-p_{31}x)-2(z_y(y)^2-p_{31}^2x^2)l(\mathbf{r}))+xy^2c_{22}+yxc_{23}+c_{24}(p_{31}x+z_y(y))\frac{l(\mathbf{r})}{p_{31}}-c_{24}x),\\
F_6^{\{1\}}(\mathbf{r}) &= \frac{y^2x}{2}.
\end{aligned}
\tag{34}
$$

The antiderivatives of $\mathrm{Q}^T(\mathbf{r})\,\mathrm{Q}(\mathbf{r})$ are

$$
F^{\{2\}}(\mathbf{r}) = \int \mathrm{Q}^T(\mathbf{r})\mathrm{Q}(\mathbf{r})\,\mathrm{d}\mathbf{r} = \int \begin{bmatrix} \mathrm{P}_x(\mathbf{r}) & & \\ \mathrm{P}_y(\mathbf{r}) & & \\ 1 & & \\ & \mathrm{P}_x(\mathbf{r}) & \\ & \mathrm{P}_y(\mathbf{r}) & \\ & 1 & \end{bmatrix} \begin{bmatrix} \mathrm{P}_x(\mathbf{r}) & & \\ \mathrm{P}_y(\mathbf{r}) & & \\ 1 & & \\ & \mathrm{P}_x(\mathbf{r}) & \\ & \mathrm{P}_y(\mathbf{r}) & \\ & 1 & \end{bmatrix}^T \mathrm{d}\mathbf{r} =
$$

$$
= \int \begin{bmatrix} \mathrm{P}_x^2(\mathbf{r}) & \mathrm{P}_x(\mathbf{r})\mathrm{P}_y(\mathbf{r}) & \mathrm{P}_x(\mathbf{r}) & & & \\ \mathrm{P}_x(\mathbf{r})\mathrm{P}_y(\mathbf{r}) & \mathrm{P}_y^2(\mathbf{r}) & \mathrm{P}_y(\mathbf{r}) & & & \\ \mathrm{P}_x(\mathbf{r}) & \mathrm{P}_y(\mathbf{r}) & 1 & & & \\ & & & \mathrm{P}_x^2(\mathbf{r}) & \mathrm{P}_x(\mathbf{r})\mathrm{P}_y(\mathbf{r}) & \mathrm{P}_x(\mathbf{r}) \\ & & & \mathrm{P}_x(\mathbf{r})\mathrm{P}_y(\mathbf{r}) & \mathrm{P}_y^2(\mathbf{r}) & \mathrm{P}_y(\mathbf{r}) \\ & & & \mathrm{P}_x(\mathbf{r}) & \mathrm{P}_y(\mathbf{r}) & 1 \end{bmatrix} \mathrm{d}\mathbf{r}.
$$

$$
\begin{aligned}
F_{11}^{\{2\}}(\mathbf{r}) &= \frac{1}{p_{31}^3}\Big(c_{31}xy-\frac{c_{32}}{2p_{32}^3}(2z_x(x)^2l(\mathbf{r})+yp_{32}(yp_{32}-2z_x(x)))-\frac{c_{33}}{p_{32}^2}(p_{32}y-z_x(x)l(\mathbf{r}))-\frac{c_{34}}{p_{32}}l(\mathbf{r})+\\
&+\frac{c_{35}}{4p_{32}^2}(p_{32}y(2z_x(x)-p_{32}y)-2(z_x(x)^2-p_{32}^2y^2)l(\mathbf{r}))+\frac{c_{36}}{p_{32}}Z(\mathbf{r})l(\mathbf{r})-c_{36}y),\\
F_{22}^{\{2\}}(\mathbf{r}) &= \frac{1}{p_{31}^3}\Big(c_{37}xy-\frac{c_{38}}{2p_{32}^3}(2z_x(x)^2l(\mathbf{r})+yp_{32}(yp_{32}-2z_x(x)))-\frac{c_{39}}{p_{32}^2}(p_{32}y-z_x(x)l(\mathbf{r}))-\frac{c_{40}}{p_{32}}l(\mathbf{r})+\\
&+\frac{c_{41}}{4p_{32}^2}(p_{32}y(2z_x(x)-p_{32}y)-2(z_x(x)^2-p_{32}^2y^2)l(\mathbf{r}))+\frac{c_{42}}{p_{32}}Z(\mathbf{r})l(\mathbf{r})-c_{42}y),\\
F_{33}^{\{2\}}(\mathbf{r}) &= xy,
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
F_{12}^{\{2\}}(\mathbf{r}) &= \frac{1}{p_{31}^3}\left(\frac{-c_{43}}{2p_{32}^3}(2z_x(x)^2 l(\mathbf{r}) + yp_{32}(yp_{32} - 2z_x(x))) - \frac{c_{44}}{p_{32}^2}(p_{32}y - z_x(x)l(\mathbf{r})) - \frac{c_{45}}{p_{32}}l(\mathbf{r}) + \right.\\
&\left. + \frac{c_{46}}{4p_{32}^2}(p_{32}y(2z_x(x) - p_{32}y) - 2(z_x(x)^2 - p_{32}^2y^2)l(\mathbf{r})) + \frac{c_{47}}{p_{32}}Z(\mathbf{r})l(\mathbf{r}) - c_{47}y + xyc_{48}\right),\\
F_{13}^{\{2\}}(\mathbf{r}) &= \frac{1}{p_{31}^2}\left(\frac{c_{25}}{4p_{32}^2}(p_{32}y(2z_x(x) - p_{32}y) - 2(z_x(x)^2 - p_{32}^2y^2)l(\mathbf{r})) + \frac{c_{26}}{p_{32}}Z(\mathbf{r})l(\mathbf{r}) - c_{26}y + c_{27}xy\right),\\
F_{23}^{\{2\}}(\mathbf{r}) &= \frac{1}{p_{31}^2}\left(\frac{c_{28}}{4p_{32}^2}(p_{32}y(2z_x(x) - p_{32}y) - 2(z_x(x)^2 - p_{32}^2y^2)l(\mathbf{r})) + \frac{c_{29}}{p_{32}}Z(\mathbf{r})l(\mathbf{r}) - c_{29}y + c_{30}xy\right),
\end{aligned}
\tag{36}
$$

$$
\begin{aligned}
&F_{44}^{\{2\}} = F_{11}^{\{2\}}, \quad F_{21}^{\{2\}} = F_{54}^{\{2\}} = F_{45}^{\{2\}} = F_{12}^{\{2\}},\\
&F_{55}^{\{2\}} = F_{22}^{\{2\}}, \quad F_{31}^{\{2\}} = F_{64}^{\{2\}} = F_{46}^{\{2\}} = F_{13}^{\{2\}},\\
&F_{66}^{\{2\}} = F_{33}^{\{2\}}, \quad F_{32}^{\{2\}} = F_{65}^{\{2\}} = F_{56}^{\{2\}} = F_{23}^{\{2\}},
\end{aligned}
\tag{37}
$$

$$
\begin{aligned}
&F_{14}^{\{2\}} = F_{15}^{\{2\}} = F_{16}^{\{2\}} = 0, \quad F_{41}^{\{2\}} = F_{42}^{\{2\}} = F_{43}^{\{2\}} = 0,\\
&F_{24}^{\{2\}} = F_{25}^{\{2\}} = F_{26}^{\{2\}} = 0, \quad F_{51}^{\{2\}} = F_{52}^{\{2\}} = F_{53}^{\{2\}} = 0,\\
&F_{34}^{\{2\}} = F_{35}^{\{2\}} = F_{36}^{\{2\}} = 0, \quad F_{61}^{\{2\}} = F_{62}^{\{2\}} = F_{63}^{\{2\}} = 0.
\end{aligned}
\tag{38}
$$

Although the coefficients K are real numbers, there might be some complex numbers showing up throughout calculations. The latter causes serious inconvenience, especially for a software implementation. These possible complex numbers are associated with the fact that function Z might be negative, and its logarithm we get in (31). But according to the constraint (12), the function Z can be either strictly positive or strictly negative. Thus, in order to get rid of any complex numbers, we can replace matrix P with matrix $-P$. This change is indeed possible since matrix P is homogenous: it defines according to (7) the projective transformation which does not change if matrix P is multiplied by any non-zero value.

Algorithm 4. *Calculation of the target coefficients for the orthotropic rectangular ROI*

Input:

- matrix $H \in \mathbb{R}^{3\times3}$ of projective normalization H,
- orthotropic rectangular ROI $R = [x_1, x_2]\times[y_1, y_2]$, $x_1 < x_2$, $y_1 < y_2$.

Output: Target coefficients $K = \mathrm{K}(H, R)$.

Step 1. Matrix $P = (p_{ij})$ is calculated: (8).
Step 2. Function Z is defined: (11).
Step 3. If $Z([x_1\ y_1]^T) < 0$, then $P := -P$ and Z is redefined.
Step 4. Coefficients c are calculated: (26), (27), (28), (29), (30).
Step 5. Functions z_x, z_y, l are defined (31).
Step 6. Antiderivative $F^{\{0\}}$ is defined: (32).
Step 7. Antiderivatives $F^{\{1\}}$ are defined: (33), (34).
Step 8. Antiderivatives $F^{\{2\}}$ are defined: (35), (36), (37), (38).
Step 9. Target coefficients K are calculated: (25).

2.6. Rectangular ROI

We have analytically calculated the target coefficients K for the orthotropic rectangular ROI above. Now we will generalize this solution for the arbitrary oriented rectangular ROI. Let us introduce the latter as an image of the rotation U of an orthotropic rectangle R_0:

$$
\begin{aligned}
&R = \mathrm{U}(R_0),\\
&R_0 = [x_1, x_2]\times[y_1, y_2],\\
&x_1 < x_2,\\
&y_1 < y_2.
\end{aligned}
$$

Let us use (19) and (20):

$$
\begin{aligned}
K^{\{0\}} &= \int_R \mathbf{r}^T\mathbf{r}\,\mathrm{d}\mathbf{r},\\
K^{\{1\}} &= \int_R \mathbf{r}^T\mathrm{Q}(\mathbf{r})\,\mathrm{d}\mathbf{r},\\
K^{\{2\}} &= \int_R \mathrm{Q}^T(\mathbf{r})\mathrm{Q}(\mathbf{r})\,\mathrm{d}\mathbf{r}.
\end{aligned}
$$

Let us introduce new coordinates $\mathbf{p} \overset{\text{def}}{=} \mathrm{U}^{-1}(\mathbf{r}) = U_2^T\mathbf{r}$, where

$$
U_2 \overset{\text{def}}{=} \begin{bmatrix} c & -s \\ +s & c \end{bmatrix}, \quad \text{where} \begin{cases} c \overset{\text{def}}{=} \cos(\alpha), \\ s \overset{\text{def}}{=} \sin(\alpha), \end{cases}
$$

then $\mathbf{r} = U_2\mathbf{p}$, which means

$$
\begin{aligned}
K^{\{0\}} &= \int_{R_0} \mathbf{p}^T U_2^T U_2\,\mathbf{p}\,\mathrm{d}\mathbf{p} = \int_{R_0} \mathbf{p}^T\mathbf{p}\,\mathrm{d}\mathbf{p},\\
K^{\{1\}} &= \int_{R_0} \mathbf{p}^T U_2^T\mathrm{Q}(U_2\mathbf{p})\,\mathrm{d}\mathbf{p},\\
K^{\{2\}} &= \int_{R_0} \mathrm{Q}^T(U_2\mathbf{p})\mathrm{Q}(U_2\mathbf{p})\,\mathrm{d}\mathbf{p}.
\end{aligned}
$$

Note that

$$
U_2^T Q = QU_6^T,
$$

$$
\text{where } U_6 \overset{\text{def}}{=} \begin{bmatrix}
c & & & -s & & \\
& c & & & -s & \\
& & c & & & -s \\
+s & & & c & & \\
& +s & & & c & \\
& & +s & & & c
\end{bmatrix},
\tag{39}
$$

Which means

$$
K^{\{1\}} = \left(\int_{R_0} \mathbf{p}^T\mathrm{Q}(U_2\mathbf{p})\,\mathrm{d}\mathbf{p}\right)U_6^T.
$$

$$\begin{aligned} & K^{\{0\}} = K_0^{\{0\}}, \\ \text{Thus } & K^{\{1\}} = K_0^{\{1\}} U_6^T, \\ & K^{\{2\}} = K_0^{\{2\}}, \end{aligned} \tag{40}$$

where

$$K_0^{\{0\}} = \int_{R_0} \mathbf{p}^T \mathbf{p}\, \mathrm{d}\mathbf{p},$$

$$K_0^{\{1\}} = \int_{R_0} \mathbf{p}^T \mathrm{Q}(U_2\mathbf{p})\, \mathrm{d}\mathbf{p},$$

$$K_0^{\{2\}} = \int_{R_0} \mathrm{Q}^T(U_2\mathbf{p})\, \mathrm{Q}(U_2\mathbf{p})\, \mathrm{d}\mathbf{p}.$$

However, the coefficients K_0 are equal to coefficients calculated for the matrix of the projective normalization $U_3^T H$ and ROI R_0:

$$K_0 = \mathrm{K}(U_3^T H, R_0),$$

where

$$U_3 \overset{\text{def}}{=} \begin{bmatrix} & U_2 & & 0 \\ & & & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} c & -s & 0 \\ +s & c & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{41}$$

is a uniform rotation matrix. Thus, the problem is reduced to the previously solved problem of the target coefficients calculation for the orthotropic rectangular ROI.

2.7. ROI consisting of rectangles

Consider the ROI which consists of rectangles:

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

Using (19):

$$K^{\{k\}} = \int_R f^{\{k\}}(\mathbf{r})\, \mathrm{d}\mathbf{r} = \sum_{i=1}^n \int_{R_i} f^{\{k\}}(\mathbf{r})\, \mathrm{d}\mathbf{r},$$

which means

$$K^{\{k\}} = \sum_{i=1}^n K_i^{\{k\}}, \tag{42}$$

where the coefficients

$$K_i^{\{k\}} = \int_{R_i} f^{\{k\}}(\mathbf{r})\, \mathrm{d}\mathbf{r}$$

can be analytically calculated via Algorithm 5: $K_i = \mathrm{K}(H, R_i)$.

3. Special cases of the affine image normalization

Aside from the affine transformation for image normalization, its special cases are widely used [44], which is usually even more computationally efficient. Let us consider sets of transformation with matrices A forming the linear manifold:

$$\mathbb{A}[S] \overset{\text{def}}{=} \left\{ A : A(\mathbf{a}), \mathbf{a} = S\mathbf{p}, \mathbf{p} = \begin{bmatrix} \mathbf{t} \\ 1 \end{bmatrix}, \mathbf{t} \in \mathbb{R}^d \right\} \subset \mathbb{R}^{2\times 3}, \tag{43}$$

where matrix $S \in \mathbb{R}^{6\times(d+1)}$ is a parameter defining the manifold, or in compact notation:

$$\mathbb{A}[S] = \left\{ A : A\left(S \begin{bmatrix} \mathbf{t} \\ 1 \end{bmatrix} \right), \mathbf{t} \in \mathbb{R}^d \right\}.$$

Algorithm 5. *Calculation of the target coefficients for the rectangular ROI*

Input:

- matrix $H \in \mathbb{R}^{3\times 3}$ of projective normalization H,
- rectangular ROI $R = \mathrm{U}(R_0)$, where $R_0 = [x_1, x_2] \times [y_1, y_2]$, $x_1 < x_2$, $y_1 < y_2$ is an orthotropic rectangle, and $\mathrm{U}(\mathbf{r}) = \begin{bmatrix} c & -s \\ +s & c \end{bmatrix} \mathbf{r}$, $\begin{cases} c = \cos(\alpha) \\ s = \sin(\alpha) \end{cases}$ is its rotation.

Output: Target coefficients $K = \mathrm{K}(H, R)$.
Step 1. Matrix U_3 is calculated: (41).
Step 2. Matrix U_6 is calculated: (39).
Step 3. $K_0 = \mathrm{K}(U_3^T H, R_0)$ is calculated via Algorithm 4.
Step 4. Target coefficients K are calculated: (40).

Algorithm 6. *Calculation of the target coefficients for the ROI consisting of rectangles*

Input:

- matrix $H \in \mathbb{R}^{3\times 3}$ of projective normalization H,
- ROI consisting of rectangles $R = \cup_{i=1}^n R_i$, $R_i \cap R_j = \varnothing, \quad i \neq j$.

Output: Target coefficients $K = \mathrm{K}(H, R)$.
Step 1. $K_i = \mathrm{K}(H, R_i)$ are calculated via Algorithm 5.
Step 2. Target coefficients K are calculated: (42).

Thus we can define sets of scaling, translation, shearing, and their superposition matrices. But we cannot introduce, for example, a set of rotation matrices. Let us provide some examples.

The isotropic scaling

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \mathbb{A}[S] = \left\{ \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_1 & 0 \end{bmatrix} \right\}. \tag{44}$$

The superposition of translation and shearing:

$$S = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \Rightarrow \mathbb{A}[S] = \left\{ \begin{bmatrix} 1 & t_1 & t_2 \\ 0 & 1 & t_3 \end{bmatrix} \right\}. \tag{45}$$

The superposition of shearing and anisotropic scaling:

$$S=\begin{bmatrix}1&0&0&0&0\\0&0&0&0&0\\0&1&0&0&0\\0&0&0&0&0\\0&0&1&0&0\\0&0&0&1&0\end{bmatrix}\Rightarrow \mathbb{A}[S]=\left\{\begin{bmatrix}t_1&0&t_2\\0&t_3&t_4\end{bmatrix}\right\}. \quad (46)$$

The full affine transformation

$$S=\begin{bmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\end{bmatrix}\Rightarrow \mathbb{A}[S]=\left\{\begin{bmatrix}t_1&t_2&t_3\\t_4&t_5&t_6\end{bmatrix}\right\}. \quad (47)$$

Now let us find a matrix from the given manifold $\mathbb{A}[S]$, which corresponds to the most accurate image normalization according to the RMS criterion:

$$A^*=\arg\min_{A\in\mathbb{A}[S]}\int_R\left\|\mathbf{r}-A\begin{bmatrix}\mathrm{P}(\mathbf{r})\\1\end{bmatrix}\right\|_2^2\mathrm{d}\mathbf{r}.$$

Following reasoning from subsection 2.3:

$$A^*=A(\mathbf{a}^*),\quad \mathbf{a}^*=\arg\min_{\mathbf{a}=S\begin{bmatrix}\mathbf{t}\\1\end{bmatrix}}\left(K^{\{0\}}-2K^{\{1\}}\mathbf{a}+\mathbf{a}^TK^{\{2\}}\mathbf{a}\right),$$

where $K=\mathrm{K}(H,R)$ is calculated via Algorithms 2, 3, or 6. Then

$$\mathbf{a}^*=S\begin{bmatrix}\mathbf{t}^*\\1\end{bmatrix}, \quad (48)$$

where

$$\mathbf{t}^*=\arg\min_{\mathbf{t}}\left(K_*^{\{0\}}-2K_*^{\{1\}}\begin{bmatrix}\mathbf{t}\\1\end{bmatrix}+\begin{bmatrix}\mathbf{t}^T&1\end{bmatrix}K_*^{\{2\}}\begin{bmatrix}\mathbf{t}\\1\end{bmatrix}\right),$$

where

$$\begin{aligned}K_*^{\{0\}}&\overset{\text{def}}{=}K^{\{0\}},\\K_*^{\{1\}}&\overset{\text{def}}{=}K^{\{1\}}S,\\K_*^{\{2\}}&\overset{\text{def}}{=}S^TK^{\{2\}}S.\end{aligned} \quad (49)$$

But

$$\begin{bmatrix}\mathbf{t}\\1\end{bmatrix}=I\,\mathbf{t}+\mathbf{i},$$

where

$$I\overset{\text{def}}{=}\begin{bmatrix}1&\cdots&0\\\vdots&\ddots&\vdots\\0&\cdots&1\\0&\cdots&0\end{bmatrix},\quad \mathbf{i}\overset{\text{def}}{=}\begin{bmatrix}0\\\vdots\\0\\1\end{bmatrix},$$

thus

$$\begin{aligned}\mathbf{t}^*&=\arg\min_{\mathbf{t}}\left(K_*^{\{0\}}-2K_*^{\{1\}}(I\mathbf{t}+\mathbf{i})+\right.\\&\left.+(I\mathbf{t}+\mathbf{i})^TK_*^{\{2\}}(I\mathbf{t}+\mathbf{i})\right)=\\&=\arg\min_{\mathbf{t}}\left(K_{**}^{\{0\}}-2K_{**}^{\{1\}}\mathbf{t}+\mathbf{t}^TK_{**}^{\{2\}}\mathbf{t}\right),\end{aligned}$$

where

$$\begin{aligned}K_{**}^{\{0\}}&\overset{\text{def}}{=}K_*^{\{0\}}-2K_*^{\{1\}}\mathbf{i}+\mathbf{i}^TK_*^{\{2\}}\mathbf{i},\\K_{**}^{\{1\}}&\overset{\text{def}}{=}(K_*^{\{1\}}-\mathbf{i}^TK_*^{\{2\}})I,\\K_{**}^{\{2\}}&\overset{\text{def}}{=}I^TK_*^{\{2\}}I,\end{aligned} \quad (50)$$

from which follows the analytical solution:

$$\mathbf{t}^*=\left(K_{**}^{\{2\}}\right)^{-1}\left(K_{**}^{\{1\}}\right)^T. \quad (51)$$

Algorithm 7. *Optimal special affine image normalization search*

Input:

- matrix $H\in\mathbb{R}^{3\times3}$ of projective normalization H,
- ROI $\mathrm{R}\subset\mathbb{R}^2$: $0<S(R)<\infty$ or $0<|R|<\infty$,
- matrix S, which defines linear manifold of $\mathbb{A}[S]$ matrices.

Output:

- matrix $A^*\in\mathbb{A}[S]$ of optimal affine approximation of H on R: (9),
- the corresponding value of RMS criterion of accuracy L_2^*: (6).

Step 1. $K=\mathrm{K}(H,R)$is calculated via Algorithms 2, 3, or 6.
Step 2. K_* is calculated: (49).
Step 3. K_{**} is calculated: (50).
Step 4. $\mathbf{t}^*$ is defined: (51).
Step 5. $\mathbf{a}^*$ is defined: (48).
Step 6. $A^*=A(\mathbf{a}^*)$ is calculated: (14).
Step 7. $L_2^*=L_2(\mathrm{A}^*,\mathrm{H};R)$ is calculated: (24).

Because of the example (47), Algorithm 7 is a generalization of Algorithm 1. Its program implementation in MatLab is available at *https://github.com/konovalenko-iitp/optimal-affine-image-normalization*.

4. Accelerated approach to image normalization

After the problem of optimal (special) affine image normalization was solved analytically for many cases, we can propose an accelerated approach to image normalization. This approach is based on the replacement of the projective normalization with the (special) affine one if there is no significant loss of accuracy.

Algorithm 8. *Accelerated image normalization*

Input:
- input image I_{input},
- matrix $H \in \mathbb{R}^{3\times 3}$ of projective normalization H,
- ROI $R \subset \mathbb{R}^2$: $0 < S(R) < \infty$ or $0 < |R| < \infty$,
- matrix S which defines linear manifold of $\mathbb{A}[S]$ matrices,
- accuracy threshold $L_2^{\max}$.

Output: Projectively or optimally affinely normalized image: I_{proj} or I^*_{affin}.

Step 1. Calculation of the (special) affine approximation A^* of H normalization and a corresponding value of RMS criterion of accuracy L_2^* via Algorithm 7.

Step 2. If $L_2^* \le L_2^{\max}$, then I^*_{affin} is calculated via the application of the transformation A^* to the image I_{input}. If otherwise, I_{proj} is calculated via the application of the transformation H to the image I_{input}.

An example of the optimal affine normalization is illustrated in Fig. 1. Each of these three images I_{input}, I_{proj} and I_{affin} has three channels of 1434×966 pixels. Computations were performed on a computer with an Intel Core i3 4030U processor. OpenCV library was utilized for image normalization. As the ROI R we have chosen the composition of three rectangles of text fields on a credit card. The analytical search of the optimal affine normalization (Algorithm 4) in this case on the average of 10^4 repetitions took $t_c = 0.191$ milliseconds. The application of the resulting affine normalization took $t_a = 5.90$ milliseconds, while the projective normalization took $t_p = 9.91$ milliseconds. Thus, the Algorithm (5) allowed for the $t_p/(t_c + t_a) \approx 1.63$ times faster performance.

And Fig. 1 shows that even though the camera optical axis is oriented significantly off the perpendicular to the document surface, text fields of the credit card were normalized with high accuracy.

Conclusion

In this work, we propose a fast approach for image normalization. It utilizes the affine normalization instead of projective if there is no significant loss of accuracy. The approach is based on a proposed criterion for the normalization accuracy: root mean square (RMS) coordinate discrepancies over the region of interest (ROI). The problem of optimal affine normalization according to this criterion is considered. We have established that this unconstrained optimization is quadratic and can be reduced to the problem of fractional quadratic functions integration over the ROI. The latter was solved analytically in the case of OCR where the ROI consists of rectangles. The proposed approach is generalized for various cases when instead of an affine transform its special cases are used: scaling, translation, shearing, and their superposition, allowing the image normalization process to be further accelerated.

References

[1] Zeynalov R, Velizhev A, Konushin A. Vosstanovlenie formy stranicy teksta dlya korrekcii geometricheskih iskazhenij [In Russian]. Proc of the 19 International Conference GraphiCon-2009 2009: 125-128.

[2] Zhukovskiy AE, Nikolaev DP, Arlazarov VV, et al. Segments graph-based approach for document capture in a smartphone video stream. ICDAR 2017; 1: 337-342. DOI: 10.1109/ICDAR.2017.63.

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

[4] Shemiakina JA, Faradjev IA, Zhukovsky AE. Research on algorithms for calculation of projective transformation in the problem of planar-object targeting by feature points. Sci Tech Inf Process 2018; 45(5): 346-351.

[5] Skoryukina N, Shemyakina J, Arlazarov VL, Faradzhev I. Document localization algorithms based on feature points and straight lines. Proc SPIE 2018; 10696: 106961H. DOI: 10.1117/12.2311478.

[6] Povolotskiy MA, Kuznetsova EG, Khanipov TM. Russian license plate segmentation based on dynamic time warping. Proc ECMS 2017: 285-291.

[7] Skoryukina NS, Chernov TS, Bulatov KB, et al. Snapscreen: TV-stream frame search with projectively distorted and noisy query. Proc SPIE 2017; 10341; 103410Y. DOI: 10.1117/12.2268735.

[8] Xie Y, Tang G, Hoff W. Geometry-based populated chessboard recognition. Proc SPIE 2018; 10696: 1069603.

[9] Arvind CS, Ritesh Mishra, Kumar Vishal, Venugopal Gundimeda. Vision based speed breaker detection for autonomous vehicle. Proc SPIE 2018; 10696: 106960E.

[10] Dubuisson M-P, Jain AAK. A modified Hausdorff distance for object matching. Proc 12th International Conference on Pattern Recognition 1994; 1: 566-568.

[11] Sim D-G, Kwon O-K, Park R-H. Object matching algorithms using robust Hausdorff distance measures. IEEE Trans Image Process 1999; 8(3): 425-429.

[12] Orrite C, Herrero JE. Shape matching of partially occluded curves invariant under projective transformation. Comput Vis Image Underst 2004; 93(1): 34-64.

[13] Nikolayev PP. Projectively invariant description of nonplanar smooth figures. 1. Preliminary analysis of the problem [In Russian]. Sensornye Sistemy 2016; 30(4): 290-311.

[14] Balitskii AM, Savchik AV, Konovalenko IA, Gafarov RF. On projectively invariant points of an oval with a distinguished exterior line. Probl Inf Transm 2017; 53(3): 279-283.

[15] Savchik AV, Nikolaev PP. Metod proektivnogo sopostavleniya dlya ovalov s dvumya otmechennymi tochkami [In Russian]. Informacionnye Tekhnologii i vychislitel'nye Sistemy 2018; 2018(1): 60-67.

[16] Katamanov SN. Avtomaticheskaya privyazka izobrazhenij geostacionarnogo sputnika MTSAT-1R [In Russian]. Sovremennye Problemy Distancionnogo Zondirovaniya Zemli iz Kosmosa 2007; 1(4): 63-68.

[17] Karpenko S, Konovalenko I, Miller A, et al. UAV Control on the basis of 3D Landmark Bearing-Only observations. Sensors 2015; 15(12): 29802-29820. DOI: 10.3390/s151229768.

[18] Kholopov I.S. Projective distortion correction algorithm at low altitude photographing. Computer Optics 2017; 41(2): 284-290. DOI: 10.18287/0134-2452-2017-41-2-284-290.

[19] Legge GE, Pelli DG, Rubin GS, et al. Psychophysics of reading. I. Normal vision. Vision Res 1985; 25(2): 239-252.
[20] Forsyth DA, Ponce J. Computer vision: a modern approach. Prentice Hall Professional Technical Reference; 2002.
[21] Triputen V, Gorohovatskij V. Algoritm parallel'noj normalizacii affinnyh preobrazovanij dlya cvetnyh izobrazhenij [In Russian]. Radioelektronika i Informatika 1997; 1: 97-98.
[22] Putyatin EP, Prokopenko DO, Pechenaya EM. Voprosy normalizacii izobrazhenij pri proektivnyh preobrazovaniyah [In Russian]. Radioelektronika i Informatika 1998; 2(3): 82-86.
[23] Wolberg G. Digital image warping. Los Alamitos, CA: IEEE Computer Society Press; 1990.
[24] Trusov A, Limonova E. The analysis of projective transformation algorithms for image recognition on mobile devices. Proc SPIE 2020; 11433: 114330Y.
[25] Gruen A. Adaptive least squares correlation: a powerful image matching technique. South African Journal of Photogrammetry, Remote Sensing and Cartography 1985; 14(3): 175-187.
[26] Ohta T-i, Maenobu K, Sakai T. Obtaining surface orientation from texels under perspective projection. IJCAI'81 1981; 2: 746-751.
[27] Pavić Darko Schönefeld V, Kobbelt L. Interactive image completion with perspective correction. Visual Comput 2006; 22(9-11): 671-681.
[28] Heckbert PS. Fundamentals of texture mapping and image warping. Technical Report. Berkeley: University of California, 1989.
[29] Lorenz H, Döllner J. Real-time piecewise perspective projections. GRAPP 2009: 147-155.
[30] Huang J-B, Singh A, Ahuja N. Single image superresolution from transformed self-exemplars. Proc IEEE Conf CVPR 2015: 5197-5206.
[31] 3D Pose from three corresponding points under weak-perspective projection. Technical Report. Cambridge, MA: Massachusetts Institute of Technology; 1992.
[32] Kutulakos KN, Vallino J. Affine object representations for calibration-free augmented reality. Proc IEEE Virtual Reality Annual International Symposium 1996: 25-36.
[33] Aradhye H, Myers GK. Method and apparatus for recognition of symbols in images of three-dimensional scenes. US Patent 7,738,706 of June 15, 2010.
[34] Mikolajczyk K, Schmid C. An affine invariant interest point detector. In Book: Heyden A, Sparr G, Nielsen M, Johansen P, eds. Computer vision – ECCV 2002. Berlin, Heidelberg, New York: Springer-Verlag; 2002: 128-142.
[35] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. Int J Comput Vis 2004; 60(1): 63-86.
[36] Morel J-M, Yu G. ASIFT: A new framework for fully affine invariant image comparison. SIAM J Imaging Sci 2009; 2(2): 438-469.
[37] Kadir T, Zisserman A, Brady M. An affine invariant salient region detector. In Book: Pajdla T, Matas J, eds. Computer vision – ECCV 2004. Berlin, Heidelberg, New York: Springer-Verlag; 2004: 228-241.
[38] Faugeras OD. What can be seen in three dimensions with an uncalibrated stereo rig? In Book: Sandini G, ed. Computer vision – ECCV'92. Berlin, Heidelberg, New York: Springer-Verlag; 1992: 563-578.
[39] Zwicker M, Räsänen J, Botsch M, et al. Perspective accurate splatting. Proceedings of Graphics Interface 2004: 247-254.
[40] Kunina IA, Gladilin SA, Nikolaev DP. Blind compensation of radial distortion in a single image using fast Hough transform. Computer Optics 2016; 40(3): 395-403. DOI: 10.18287/2412-6179-2016-40-3-395-403.
[41] Hsu SC, Sawhney HS. Influence of global constraints and lens distortion on pose and appearance recovery from a purely rotating camera. Proc 4th IEEE Workshop on Applications of Computer Vision (WACV'98) 1998: 154-159.
[42] Chen H, Sukthankar R, Wallace G, Li K. Scalable alignment of large-format multi-projector displays using camera homography trees. IEEE Visualization (VIS 2002) 2002: 339-346.
[43] Konovalenko IA, Kokhan VV, Nikolaev DP. Optimal affine approximation of image projective transformation [In Russian]. Sensornye Sistemy 2019; 33(1): 7-14.
[44] Vanichev AY. Normalizaciya siluetov ob"ektov v sistemah tekhnicheskogo zreniya [In Russian]. Programmnye Produkty i Sistemy 2007; 3: 86-88.

Authors' information

Ivan Andreevich Konovalenko is a researcher at the IITP RAS, a researcher-programmer at Smart Engines Service LLC. He graduated from MIPT in 2014. His major research interests include Computer Vision, Applied Mathematics, and Mathematical Analysis. E-mail: *konovalenko@smartengines.com* .

The information about author **Vladislav Vladimirovich Kokhan** you can find on page 77 of this issue.

Dmitry Petrovich Nikolaev, Ph. D. in Physics and Mathematics, is a head of the laboratory at the IITP RAS, a technical director of Smart Engines Service LLC. He graduated from Lomonosov MSU in 2000. His major research interests include Machine Vision, Algorithms for Fast Image Processing, Pattern Recognition. E-mail: *dimonstr@iitp.ru* .

Received May 25, 2020. The final version – September 28, 2020.