Log-polar Transformation as a Tool for Text Skew Estimation

D. Brodic[1], Z. N. Milivojevic[2]
[1]*University of Belgrade, Technical Faculty in Bor,*
V.J.12, 19210 Bor, Serbia, phone: +381 30 424 555
[2]*Technical College Niš,*
Aleksandra Medvedeva 20, 18000 Niš, Serbia, phone: +381 18 588 211
dbrodic@tf.bor.ac.rs

Abstract—**The paper proposed the method for text skew detection based on log-polar transformation. The original image is transformed in the log-polar domain as well as the control ellipse. Theirs cross-correlation established the cost function. The extraction of the cost function maximums gives the text skew value in the left and right region from the centre point of transformation. The method is suitable for the printed text. It is characterized by the accuracy and computational time inexpensiveness.**

Index Terms—**Document image processing, text skew, log-polar transformation.**

I. INTRODUCTION

The printed text is characterized with an articulated regularity in shape [1]. It means that the letters have the similar sizes. The distance between text lines is generally adequate. The inter line spacing is sufficient to split text lines. The orientation of the text lines is look-alike. It forms uniform text skew. All listed attributes represent relatively predicted characteristics. They simplify the skew estimation of the printed text. It should be noted that the text skew occurrence is unavoidable. It is an implication of the digitization process. However, its existence could cause the optical character recognition system (OCR) failing. Hence, the text skew estimation represents the crucial step in OCR [2].

This paper introduces a new algorithm for text skew estimation. It is based on the co-operation of the two methods, log-polar transformation and cross-correlation. First, it converts two images, i.e. the original image and the image with referent text skew, in log-polar space. Furthermore, both images are cross-correlated in the log-polar domain in order to extract theirs similarity. As a result, the cross-correlation function has been obtained. This cost function represents the similarity measure between the images [3]. Hence, its maximum values give the optimum matching. The function's horizontal axis represents the angle of the image rotation. This way, its maximums identify the text skew of the image. The proposed algorithm shows a good skew estimation in the standard image resolution of

300 dpi. It is computationally inexpensive, too. This is a consequence of the binary image processing. Hence, this method is a promising one.

Organization of this paper is as follows. Section 2 describes briefly the principle of the log-polar transformation. Section 3 proposes the algorithm for the estimation of the text skew based on log-polar transformation. Section 4 explains the steps of the algorithm. Section 5 defines text experiments. Section 6 discusses obtained results. Section 7 makes conclusion.

II. LOG-POLAR TRANSFORMATION

Log-polar transformation maps points from the image, i.e. Cartesian space to the points in the log-polar parametric space. It reduces data vector depending on position the in the image. Consequently, it is based on the eye retina of the primates [3]. First, Cartesian space coordinates x and y are converted into polar coordinates radius r and angle θ. Their mapping is as follows [3]:

$$r = \sqrt{x^2 + y^2}, \tag{1}$$

$$\theta = \arctan\frac{y}{x}. \tag{2}$$

Furthermore, the log transformation is obtained as

$$\rho = \ln r = \ln\sqrt{x^2 + y^2}. \tag{3}$$

Hence, the (x, y) coordinates from Cartesian space are mapped to (ρ, θ) coordinates in log-polar space.

III. ALGORITHM

Document text image represents the image, which is a product of the image scanning process. It is a digital grey-level image given by matrix **D**. It consists of M rows, N columns, and L intensity levels of gray. L is the integer from $\{0,\ldots,255\}$. Hence, $D(i, j) \in \{0,\ldots,255\}$, where $i = 1,\ldots,M$ and $j = 1,\ldots,N$. After the binarization process, the image is transformed into binary one, i.e. **B**. Its elements are equal to 0 or 1. If $D(i, j) \geq D_{th}(i, j)$ then $B(i, j) = 1$, else if $D(i, j) < D_{th}(i, j)$ then $B(i, j) = 0$. Accordingly, $D_{th}(i, j)$ represents the global threshold sensitivity decision value [4], [5].

Currently, document text image is given as binary matrix B featuring M rows and N columns, and two levels of intensity: 0 and 1.

Log-polar transformation is a nonlinear and non-uniform sampling of spatial domain. Nonlinearity has been introduced by polar mapping, while non-uniform sampling is the result of logarithmic scaling [6]. Consider the log-polar coordinate system, which denotes the radial distance and the angle from the centre. For the input binary image **B**, the centre point of transformation has been extracted as $B(m_c, n_c)$. The radius is assigned as R. It ensures the maximum number of pixels that have to be included within the reference circle of the conversion. Centre of the circle is given as $m_c = M/2$, and $n_c = N/2$ [7]. Furthermore, the image is converted into the polar coordinate system. This way, the binary image **B** has been transformed into the polar domain (r, θ), where [6]:

$$r = \sqrt{(i - m_c)^2 + (j - n_c)^2}, \ 0 \le r \le R, \qquad (4)$$

$$\theta = \arctan\left(\frac{j - n_c}{i - m_c}\right), \ 0° \le \theta \le 360°, \qquad (5)$$

and $i = 1, …, M, j = 1, …, N$. Furthermore, log-polar transform is given as (ρ, θ) where ρ is obtained from (3). In the log-polar domain matrices of text image **B** and referent object **E** are marked as **BC** and **EC**. Cross-correlation is a similarity measure between two images. Its function is

$$CC(\theta) = C_{coeff}\left(\mathbf{BC}, circshift(\mathbf{EC}, \theta)\right), \qquad (6)$$

where **ECS** is $circshift(\mathbf{EC}, \theta)$ and $C_{coeff}(\mathbf{BC}, \mathbf{ECS})$ is given as

$$C_{coeff} = \frac{\sum_\rho \sum_\theta (BC_{\rho\theta} - \overline{BC})(ECS_{\rho\theta} - \overline{ECS})}{\sqrt{\left(\sum_\rho \sum_\theta (BC_{\rho\theta} - \overline{BC})^2\right)\left(\sum_\rho \sum_\theta (ECS_{\rho\theta} - \overline{ECS})^2\right)}}. \qquad (7)$$

If the images are more alike, then cross-correlation function $CC(\theta)$ will tend to approach 1.

The identification of the rotation in spatial domain, i.e. image space is a complex task. However, the rotation in log-polar space is mapped into translation. Translation in direction of one axis is an easy task to solve. Suppose that a referent object is rotated in a space domain. If it is cross-correlated with text image for different angles, then it will be readout as translation in log-polar space. First, the objective is the selection of the correct referent object. The right choice can be an ellipse. It is a suitable object because it can overlap text efficiently. However, the ellipse has to be normalized according to the text image dimension. Furthermore, it is split into left and right part from the centre point of the transformation. This way, parts of the ellipse are matched with the original image by cross-correlation. Hence, they establish left and right skew estimation. Unlike the other methods, proposed algorithm identifies two skews: left and right one. This fact is the advantage of it.

IV. ALGORITHM PROCEDURE

The algorithm for text skew estimation based on the log-polar transformation consists of the following steps:

1) Text image extraction by a bounding box (text image).
2) Identification of the centre point for log-polar transformation.
3) Creation of the binary image with the normalized ellipse (ellipse image).
4) Log-polar transformation of text image.
5) Log-polar transformation of ellipse image.
6) Cross-correlation between text and ellipse image in the log-polar domain.
7) Extraction of the maximum values from cross-correlation cost function.
8) Identification of left and right side skew angle from the centre transformation point.

Step 1. The original binary text image **B** extracted by bounding box is shown in Fig. 1.
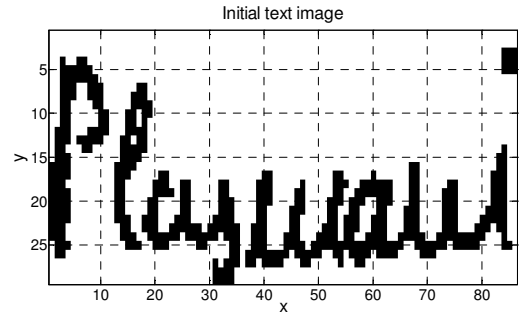

Fig. 1. Original binary text image.

Step 2. Transformation centre point $B(m, n)$ is extracted according to the pixel density at the centre of the bounding box.

Step 3. The ellipse is created as the binary image. The size of the ellipse depends on the original text image, which means that it is normalized. The ellipse is shown in Fig. 2.
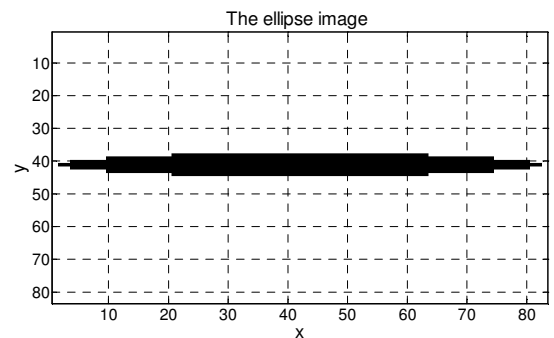

Fig. 2. Normalized ellipse.

From (4)–(6), log-polar transformation of the text and ellipse image has been achieved.

Step 4. Fig. 3 shows log-polar transformation of the text image.

Step 5. Fig. 4 shows log-polar transformation of the ellipse.

Step 6. The cross-correlation of the text and ellipse image in the log-polar domain results with the cross-correlation cost function. It is shown in Fig. 5.
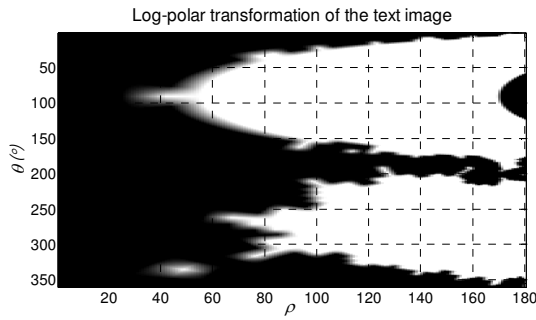
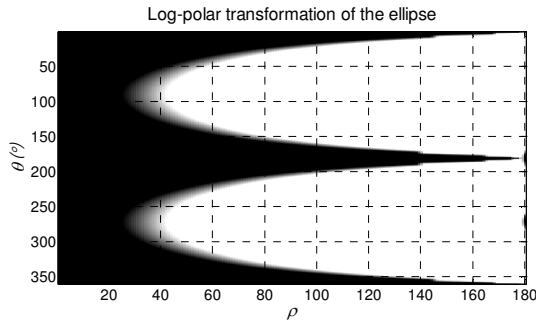Fig. 3. Log-polar transformation of text image.

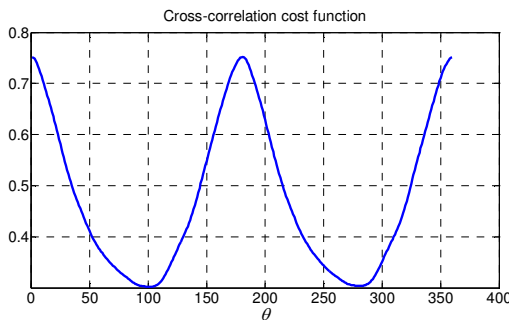Fig. 4. Log-polar transformation of ellipse image.
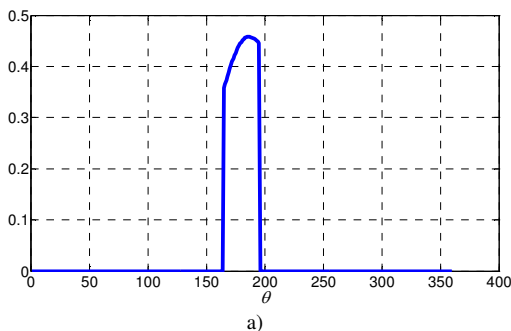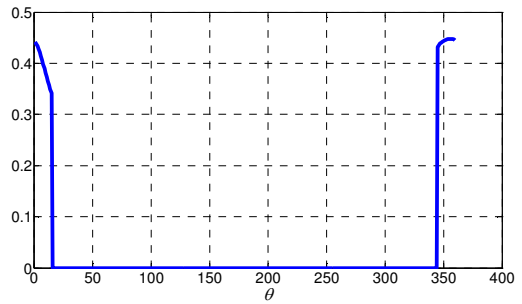

Fig. 5. Cross-correlation cost function.

Step 7. From Fig. 5, cost function has two maximums (around 0° and 180°). These maximums represent two angles of the text skew given from the transformation centre point (left and right). Furthermore, information has been returned from the log-polar domain to the spatial image domain. Excerpt from the cross-correlation cost function for left and right part from the transformation point is shown in Fig. 6.

Step 8. As a result, left and right skew angle lines extracted from the cost function are drawn. Fig. 7 shows these lines.

a)

b)

Fig. 6. Cross-correlation cost function: for the left ellipse (a), and right ellipse (b) with text image.
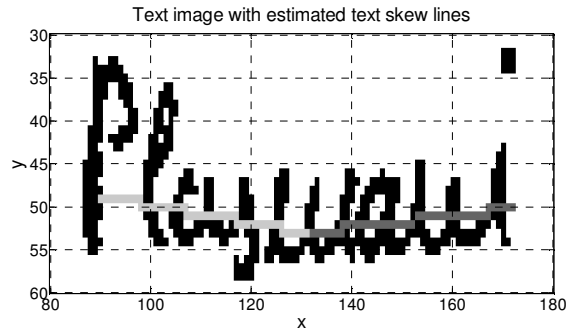

Fig. 7. Enlarged image with skew estimation lines.

V. EXPERIMENTS

The main task of the experiments is the evaluation of the algorithm's ability to estimate correctly the text skew. Experiments were performed mostly on synthetic data sets. They represent single-line samples of the printed text [9]. The text is rotated for the angle θ, from 0° to 60° by 5° steps around x-axis. It is shown in Fig. 8.
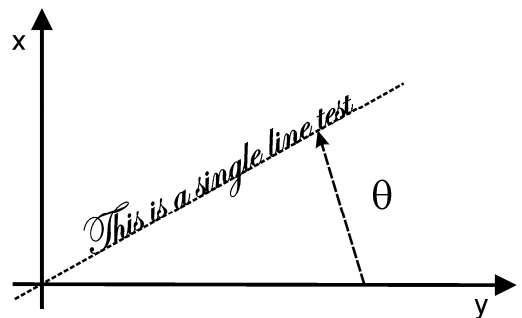

Fig. 8. A printed text sample which is rotated up to 60°.

Furthermore, all text samples are given in the standard resolution of 300 dpi. The results are evaluated by absolute deviation, i.e. error. It is given as:

$$\Delta\theta_A = \left|\theta_{REF} - \theta_A\right|, \qquad (8)$$

where θ_{REF} is referent skew of the input text sample and θ, i.e. θ_A is skew of the text sample obtained with a tested algorithm. Moreover, the relative error (*RE*) [9] is an important measure as well. The reference line hit rate, i.e. *RLHR* incorporates it. It is defined as [8], [9]

$$RLHR = 1 - RE(\theta) = 1 - \frac{\Delta\theta_A}{\theta_{REF}}, \qquad (9)$$

where $RE(\theta)$ represents a relative error.

VI. RESULTS AND DISCUSSION

The result of testing is given in Table I in Appendix. The algorithm shows good results in the whole range of testing angles. Hence, proposed method is promising in the domain of the text skew accuracy. Furthermore, it is computer time non-intensive.

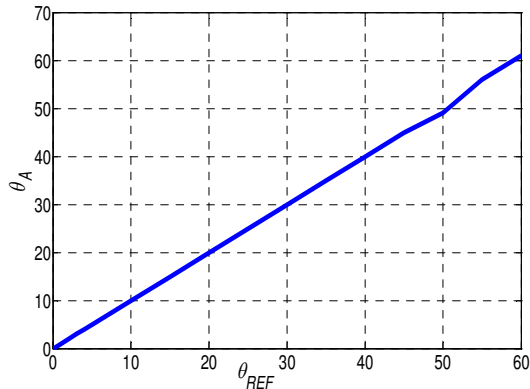Fig. 9 shows θ_A as a linear function of θ_{REF}.

Fig. 9. θ_A as an almost linear function of θ_{REF}.

Fig. 10 shows *RLHR*. It has a very good value in the whole range of testing angles.
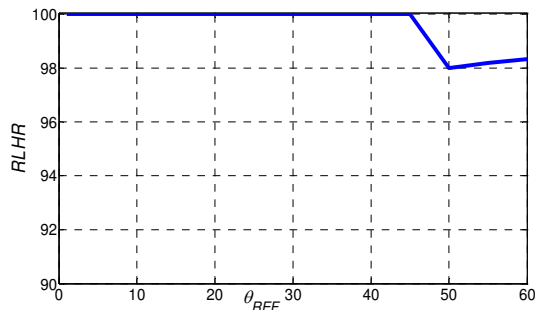

Fig. 10. *RLHR* (Reference line hit rate).

VII. CONCLUSIONS

The paper introduced the method for text skew estimation based on log-polar transformation and cross-correlation. It estimates the similarity between text image and ellipse in the log-polar domain. As a result, the cross-correlation cost function is obtained. Its maximums give text skew left and right from the transformation centre point. Proposed method shows good results for skew estimation of the printed and hand printed text. The future investigation will exploit the estimation of skew for the handwritten text.

APPENDIX A

TABLE I. TESTING RESULTS.

θ_{REF} (º)	θ_A (º)	$\Delta\theta_A$ (º)	$RE(\theta)$	$RLHR$ (%)
0	0	0	-	-
1	1	0	0.000	100.00
2	2	0	0.000	100.00
3	3	0	0.000	100.00
4	4	0	0.000	100.00
5	5	0	0.000	100.00
10	10	0	0.000	100.00
15	15	0	0.000	100.00
20	20	0	0.000	100.00
25	25	0	0.000	100.00
30	30	0	0.000	100.00
35	35	0	0.000	100.00
40	40	0	0.000	100.00
45	45	0	0.000	100.00
50	49	1	0.020	98.00
55	56	1	0.018	98.18
60	61	1	0.016	98.33

REFERENCES

[1] D. Brodic, "The Evaluation of the Initial Skew Rate for Printed Text", "Journal of Electrical Engineering - Elektrotechnický časopis", Versita, no. 3, pp. 142–148, 2011.

[2] P. Shivakumara, et al., "A Novel Technique for Estimation of Skew in Binary Text Document Images based on Linear Regression Analysis", *Sādhanā*, Springer, no. 1, pp. 69–86, 2005.

[3] V. J. Traver, A. Bernardino, "A Review of Log-polar Imaging for Visual Perception in Robotics", *Robotics and Autonomous Systems,* Elsevier, vol. 4, pp. 378–398, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2009.10.002

[4] L. Sauvola, M. Pietikainen, "Adaptive Document Image Binarization", *Pattern Recognition*, Elsevier, no. 2, pp. 225–236, 2000.

[5] A. Khasman, B. Sekeroglu, "Document Image Binarisation Using a Supervised Neural Network", *International Journal of Neural Systems,* no. 5, pp. 405–418, 2008. [Online]. Available: http://dx.doi.org/10.1142/S0129065708001671

[6] S. Zokai, G. Wolberg, "Image Registration Using Log-Polar Mappings for Recovery of Large-Scale Similarity and Projective Transformations", *IEEE Transaction on Image Processing,* no. 10, pp. 1422–1434, 2005. [Online]. Available: http://dx.doi.org/10.1109/TIP.2005.854501

[7] M. K. Bhowmik, et al., "Classification of Log-Polar-Visual Eigenfaces using Multilayer Perceptron", *International Journal of Image Processing,* no. 1, pp. 12–23, 2010.

[8] D. Brodić, D. R. Milivojević, Z. Milivojević, 'Basic Test Framework for the Evaluation of Text Line Segmentation and Text Parameter Extraction", *Sensors*, MDPI, no. 5, pp. 5263–5279, 2010.

[9] V. S. Popov, "Principle of Symmetry and Relative Errors of Instrumentation and Transducers", *Automation and Remote Control,* no. 5, pp. 183–189, 2001.