Text Skew Detection Using Log-polar Transformation

Darko Brodic[1], Zoran N. Milivojevic[2], Dragan R. Milivojevic[3]
[1]University of Belgrade, Technical Faculty in Bor
V.J. 12, 19210 Bor, Serbia
{dbrodic@tf.bor.ac.rs}

[2]Technical College Niš, Aleksandra Medvedeva 20
18000 Niš, Serbia
{zoran.milivojevic@jotel.co.rs}

[3]Institute for Mining and Mettalurgy
Zeleni bulevar bb, 19210 Bor
Serbia

ABSTRACT: *The paper proposes the method for text skew detection based on log-polar transformation and crosscorrelation. The text image is transformed into log-polar domain as well as the control ellipse. Theirs cross-correlation established the cost function. The extraction of the cost function maximum represents the text skew value in the region. The method is characterized by the accuracy and computational time inexpensiveness.*

Keywords: Document Image Processing, Log-Polar Transformation, Text Skew

Received: 29 May 2023, Revised 29 July 2023, Accepted 5 August 2023

DOI: 10.6025/jcl/2023/14/4/113-120

1. Introduction

The printed text is a strongly formed text type with a articulated regularity in shape [1]. Accordingly, the letters are of the similar size and the distance between text lines is generally sufficient. Hence, the spacing between text lines is decent. The orientation of the text lines is similar, which leads to the uniform text skew. These attributes represent the relatively predicted characteristics, which simplify the printed skew identification.

However, the text skew extraction represents a severe problem. It is a consequence of the digitization process. Hence, the text skew occurrence is simply unavoidable. The existence of this phenomena could cause the optical character recognition system failing. Hence, its identification represents one of the crucial steps [2].

Existing methods for the text skew identification can be grouped as follows [3]: projection profiles method, k-nearest neighbor clustering method, Hough transforms method, Radon transforms method, Fourier transformation method, cross-correlation method, and other methods.

This paper recommends a new algorithm based on the interaction of the log-polar transformation and crosscorrelation. Firstly, it converts images into log-polar space. Furthermore, two images are cross-correlated in log-polar domain to extract theirs similarity. As a consequence, crosscorrelation function called cost function has been obtained. Its maximum values represent the angle of the text skew estimation. The result gives the fractured line with two text skew values: left and right. These two values represent the new elements in the skew estimation compared to the previously mentioned methods like Hough transform, Radon transform, vertical projection profiles, etc. The proposed algorithm shows a good skew estimation in the standard resolution. Hence, it contributes to theirs robustness.

Organization of this paper is as follows. Section 2 describes the proposed algorithms for the estimation of the text skew. Section 3 defines text experiments. Section 4 compares and discusses obtained results. Section 5 makes conclusions.

2. Algorithm

2.1. Document Image

Document text image is a product of the image scanning. It is a digital gray-level image, which is represented by matrix **D**. It consists of M rows, N columns, and contains the elements which intensity has L discrete levels of gray. L is the integer from $\{0, \ldots, 255\}$, $D(i, j) \in \{0, \ldots, 255\}$, where $i = 1,\ldots, M$ and $j = 1,\ldots, N$. After performing the binarization procedure the image represented by matrix **D** is transformed into binary image $B(i, j)$. Its elements are equal to 1 if $D(i, j) < D_{th}(i, j)$, or to 0 if $D(i, j) < D_{th}(i, j)$, where D_{th} is given by any local binarization method [4]–[5]. D_{th} represents local threshold sensitivity decision value. Currently, document image is given as binary matrix **B** featuring M rows and N columns.

2.2. Log-polar Transformation

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

$$r = \sqrt{(i-m)^2 + (j-n)^2}\ , 0 \le r \le R\,, \tag{1}$$

and

$$\theta = \arctan\left(\frac{j-n}{i-m}\right), 0° \le \theta \le 360°\,. \tag{2}$$

Furthermore, log-polar transform is given as (ρ, θ) where:

$$\rho = ln r\,. \tag{3}$$

Applying a polar coordinate transformation to an image maps radial lines in Cartesian space to horizontal lines in the polar coordinate space.

2.3. Cross-correlation

Cross-correlation is a measure of similarity of two images. In the discrete form, it is given as [7]:

$$cc(i, j) = B(i, j) \circ E(i, j) = \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} B(k,l)E(i+k, j+l)\,. \tag{4}$$

However, in our case cross-correlation have to be made between the images in log-polar domain. Hence, equation (4) should be transformed adequately. Hence, suppose that in the log-polar domain matrices of text image ***B*** and referent object ***E*** are marked as ***BC*** and ***EC***. Furthermore, *cc* (in log-polar domain) that represents the cross-correlation function can be defined as:

$$cc(\theta) = C_{coeff}\left(\mathbf{BC}, circshift(\mathbf{EC}, \theta)\right), \tag{5}$$

where ***ECS*** is *circshift* (***EC***, θ) and *Ccoeff* (**BC**, *circshift* (***EC***, θ)) is given as [7]:

$$C_{coeff} = \frac{\sum_{\rho}\sum_{\theta}(BC_{\rho\theta} - \overline{BC})(ECS_{\rho\theta} - \overline{ECS})}{\sqrt{\left(\sum_{\rho}\sum_{\theta}(BC_{\rho\theta} - \overline{BC})^2\right)\left(\sum_{\rho}\sum_{\theta}(ECS_{\rho\theta} - \overline{ECS})^2\right)}}, \tag{6}$$

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

The identification of the rotation in the spatial domain, i.e image space is a complex task. However, the rotation in the log-polar space is mapped into translation. The translation in the direction of one axis is an easy task to solve. Suppose that a referent object is rotated in the space domain. If it is crosscorrelated with the text image for the different angles, then it will be readout as the translation in the log-polar space. The objective is the selection of the referent object. In this paper, the ellipse is selected as a referent object. 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, the ellipse is split into left and right half part from the center point of the transformation. This way, those parts of the ellipse are matching with the original image by the cross-correlation. Hence, they establish the left and right skew estimation. Unlike the other methods, the log polar transformation identifies two skews: left and right one. This fact is the advantage of the proposed method.

2.4. Algorithm's Steps

The algorithm for the estimation of the text skew based on log-polar transformation is as follows:

1. Text image extraction by the bounding box (text image).

2. Identification of the center point needed for the logpolar transformation.

3. Creation of the binary image with normalized ellipse (ellipse image).

4. Log-polar transformation of the text image.

5. Log-polar transformation of the ellipse image.

6. Cross-correlation of the text image with ellipse image in the log-polar domain.

7. Extraction of the maximum values from the crosscorrelation function.

8. Identification of the left and right side skew angle from the center transformation point.

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

Step 2.
Typical center point $B(m, n)$ needed for transformation is extracted according to the pixel density in the center of the bounding box.

Step 3.
Furthermore, the binary image with ellipse is created. The size of the ellipse depends on the original text image. It can be said that ellipse is normalized according to the text image size. The binary image with ellipse is shown in Figure 2.

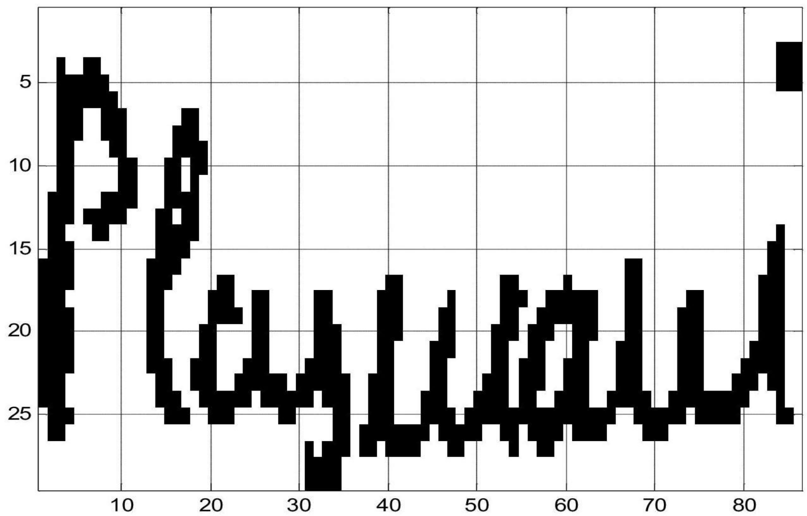

Figure 1. The original binary text image

The text is extracted by the bounding box

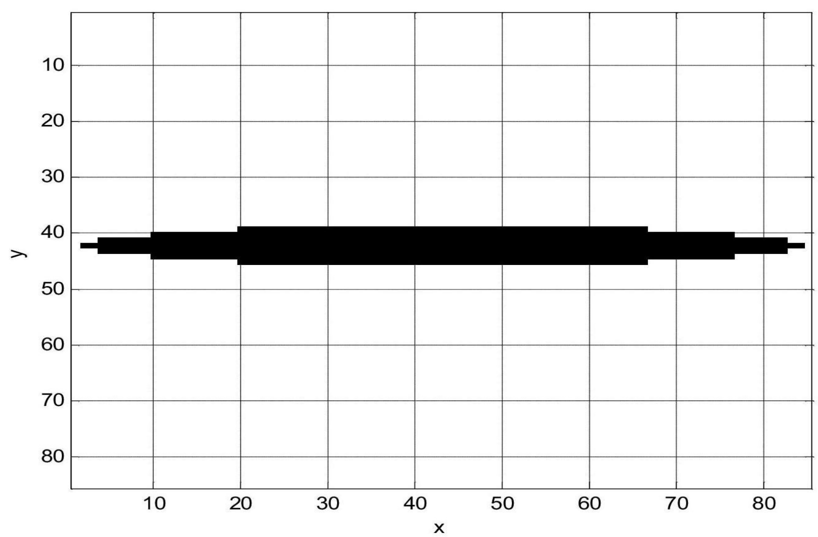

Figure 2. Ellipse image (ellipse has been normalized according to the text object)

According to the eq. (1)-(3) log-polar transformation of the text and ellipse image is achieved.

Step 4.
In Figure 3, the log-polar transformation of the text image is shown.

Step 5.
In Figure 4, the log-polar transformation of the ellipse image is shown.

Step 6.
As a result of the cross-correlation of the text and ellipse image in log-polar domain, a cross-correlation function is obtained. It is so-called cost function. This function is shown in Figure 5.

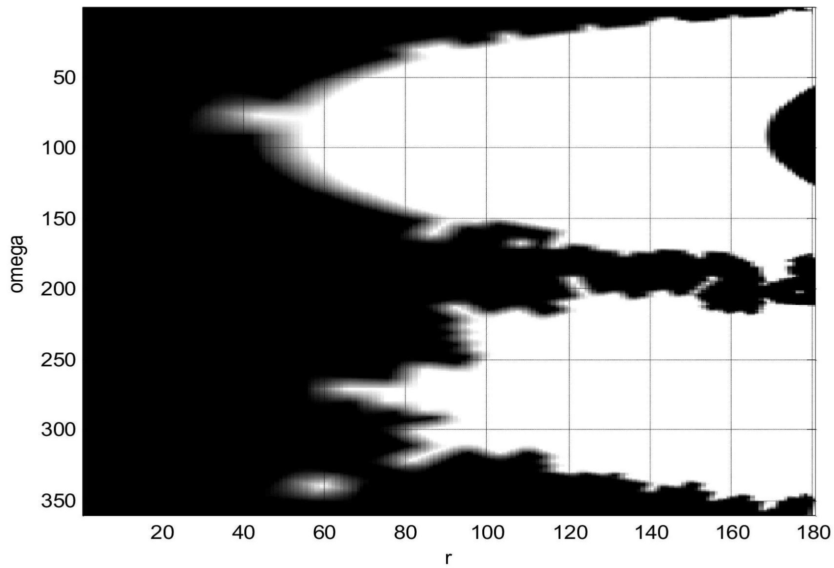

Figure 3. Log-polar transformation of the original text image

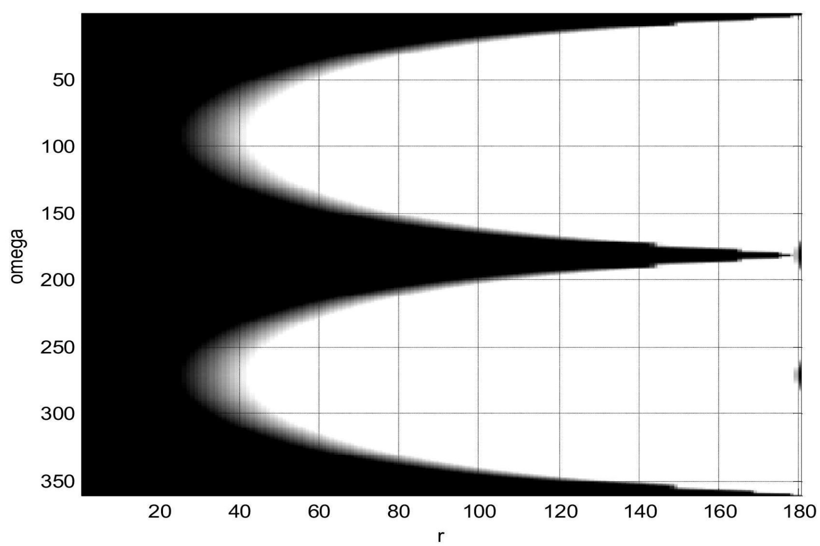

Figure 4. Log-polar transformation of the ellipse.

Step 7.
From Figure 5 the cost function has two maximums. These maximums represent the two angles of the text skew rate given from the central point of the transformation. These information has been return from the log-polar domain to the spatial image domain.

Step 8.
As a result the left and right angle skew line in drawn in the image. This is shown in the Figure 6.

3. Experiments

The main goal of the experiments is the evaluation of the algorithm for the text skew estimation. It evaluates the algorithm's performance in the skew tracking domain.

Experiments were performed mostly on the synthetic datasets, which represents the single line of the printed text sample [8].

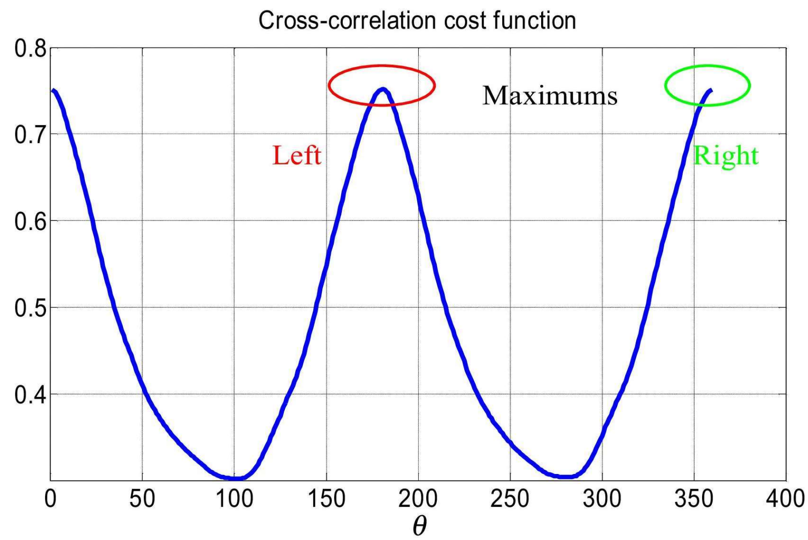

Figure 5. Cross-correlation of the original text image and ellipse in log polar domain.

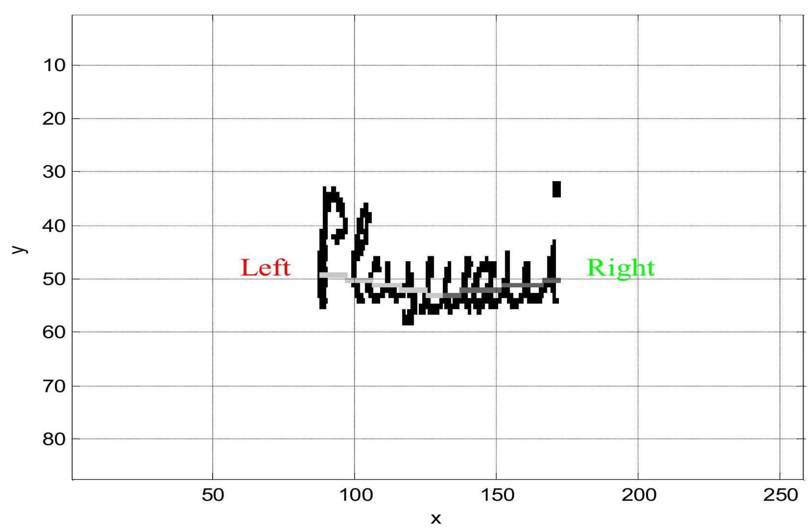

Figure 6. Enlarged original image with the skew line obtained from the log-polar cross-correlation function.

The test consists of the single line printed text rotated for the angle β from 0° to 60° by the 5° steps around *x*-axis [8]. Text sample is shown in Figure 7.

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

$$\Delta\theta_A = \left|\theta_A - \theta_{REF}\right|, \tag{7}$$

where θ_{REF} is the referent skew of the input text sample and θ i.e. θ_A is the skew of the text sample obtained with a tested algorithms. Furthermore, a relative error (*RE*) [9] is important for the algorithm evaluation as well. It is given as:

$$RE(\theta_A) = \frac{\Delta\theta_A}{\theta_{REF}} = \frac{|\theta_A - \theta_{REF}|}{\theta_{REF}}. \quad (8)$$

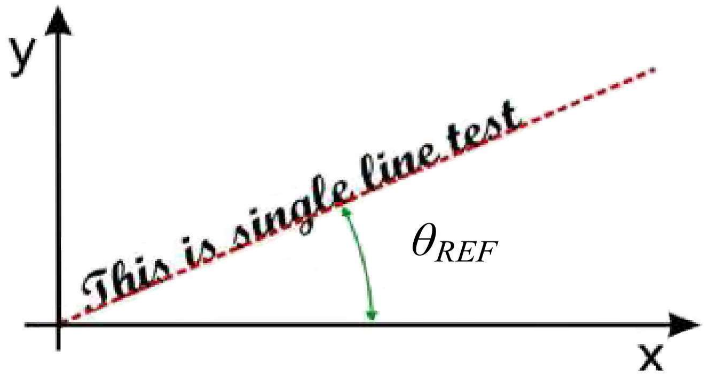

Figure 7. Enlarged original image with the skew line obtained from the log-polar cross-correlation function

4. Results and Discussion

The result of testing is given in Table 1.

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

Table 1. Testing Results

The investigated algorithm shows good results in the whole testing angle range. Hence, the presented method is promising in the domain of the accuracy. Furthermore, it is computer time non-intensive.

5. Conclusion

This paper gives the analysis of the text skew estimation techniques based on the log-polar transformation method. This method estimates the similarity of the text image and ellipse in the log-polar domain. As a result, the cross correlation cost function is obtained. Given method shows good results for the skew estimation of the printed and hand printed text. Hence, it proves its accuracy in the standard resolution of text images. Furthermore, the method is computer time inexpensive.

The future investigation will be toward the estimation of the handwritten text skew in the unconstrained text.

Acknowledgement

This work is partly supported by the project funded by the Ministry of Education and Science of the Republic of Serbia, No. TR33037.

References

[1] Brodic, D. (2011). The Evaluation of the Initial Skew Rate for Printed Text. *Journal of Electrical Engineering - Elektrotechnický asopis, 62*(3), 142-148.

[2] Shivakumara, P., Kumar, G. H., Guru, D. S., Nagabhushan, P. (2005). A Novel Technique for Estimation of Skew in Binary Text Document Images based on Linear Regression Analysis. *Sdhan, 30*(1), 69–86.

[3] Amin, A., Wu, S. (2005). Robust Skew Detection in Mixed Text/Graphics Documents. In *Proceedings of the 8th ICDAR ’05*, Seoul, Korea, 1, 247–251.

[4] Sauvola, L., Pietikainen, M. (2000). Adaptive Document Image Binarization. *Pattern Recognition, 33*(2), 225-236.

[5] Khasman, A., Sekeroglu, B. (2008). Document Image Binarisation Using a Supervised Neural Network. *International Journal of Neural Systems, 18*(5), 405-418.

[6] Bhowmik, M. K., Bhattacharjee, D., Nasipuri, M., Kundu, M., Basu, D. K. (2010). Classification of Log-Polar-Visual Eigenfaces using Multilayer Perceptron. *International Journal of Image Processing (IJIP), 4*(1), 12-23.

[7] Gonzalez, R. C., Woods, R. E. (2002). *Digital Image Processing*, 2nd edn. – Prentice-Hall.

[8] Brodi, D., Milivojevi, D. R., Milivojevi, Z. (2010). Basic Test Framework for the Evaluation of Text Line Segmentation and Text Parameter Extraction. *Sensors, 10*(5), 5263– 5279.

[9] Popov, V. S. (2001). Principle of Symmetry and Relative Errors of Instrumentation and Transducers. *Automation and Remote Control, 62*(5), 183–189.