Comparison of Different Methods for the Estimation of Text Skew

Darko Brodić [1], Zoran N. Milivojević [2], Ivo R. Draganov [3] and Viša Tasić [4]

Abstract – **The paper analyzes different methods for the evaluation of the text skew. It incorporates the widely used vertical projection method as well as the moment based and log-polar cross-correlation method. The comparison among methods is based on the printed text samples. The test consists of the document image samples in the standard resolution of 300 dpi. The analyzed algorithms show different skew accuracy. However, the method with the smallest accuracy deviation proved the advantage over the other methods. Furthermore, this contributes to its robustness in applications.**

Keywords – **document image analysis, horizontal projection profiles, initial skew rate, moments, skew estimation.**

I. Introduction

The identification of the object skew in the image is one of the most important tasks in digital image processing. Accordingly, the text skew estimation is a key step in the document image analysis, too. Its existence could cause the optical character recognition system (OCR) failing. It is due to the system sensitivity to any skew appearance in the text. Hence, the text skew estimation represents the crucial step in OCR [1].

The printed text is characterized with the regularity in shape [1]. It means that the letters have the similar sizes and the distance between text lines are adequate. It enables that between line spacing is sufficient to split text lines. Furthermore, the orientation of the text lines is similar, which forms the uniform text skew. All above attributes represent relatively predicted characteristics. Hence, they simplify the procedure of the skew estimation for the printed text.

It should be noted that the text skew occurrence is unavoidable. It is an implication of the digitization process. Existing methods for the estimation of the text skew can be classified as [1]: (i) Projection profiles method, (ii) K-nearest neighbor clustering method, (iii) Hough transforms method, (iv) Fourier transformation method, (v) Cross-correlation method, and (vi) Other methods.

[1]Darko Brodić is with the Technical Faculty in Bor, University of Belgrade, Vojske Jugoslavije 12, 19210 Bor, Serbia, E-mail: dbrodic@tf.bor.ac.rs, mjevtic@tf.bor.ac.rs.

[2]Zoran N. Milivojević is with College of Applied Technical Sciences, Aleksandra Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.milivojevic@vtsnis.edu.rs.

[3]Ivo R. Draganov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: idraganov@tu-sofia.bg

[3]Viša Tasić is with the Mining and Metalurgy Institute, Department of Industrial Informatics, Zeleni Bulevar 35, 19210 Bor, Serbia, E-mail: visa.tasic@irmbor.co.rs

In this paper, we decided to analyze widespread technique for text skew estimation based on a projection profiles calculation along with moment based method and newly proposed log-polar cross-correlation method. Furthermore, their estimation of the detected text skew is analyzed and compared.

Organization of this paper is as follows. Section 2 describes different methods for text skew detection. Section 3 defines the experiments. Section 4 presents and analyzes the results. Section 5 gives the conclusions.

II. Proposed Methods

Below is presented each of tested algorithm for text skew detection. They are: (i) horizontal projection profiles, (ii) moment based algorithm, and (iii) log-polar cross-correlation algorithm.

Projection profile method

The horizontal projection profile method extracts features from the projection profiles of text lines. Horizontal profile gives the sum of the black pixels perpendicular to the *x* axis. It is represented by the vector P_h of size N defined by:

$$P_h[j] = \sum_{i=1}^{M} B(i, j) \tag{1}$$

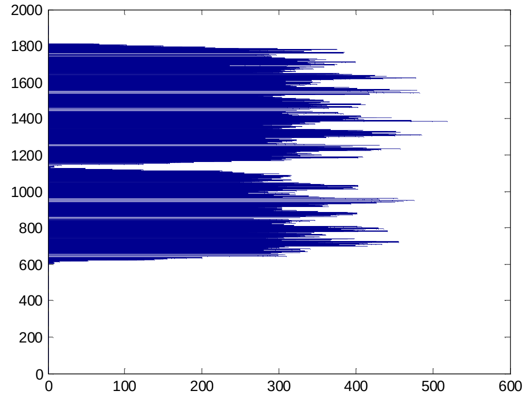
where **B** is the binary image featuring $i = 1, ..., M$ rows and $j = 1, ..., N$ columns. This vector is called a density histogram of horizontal black pixels.

Consequently, the valleys of the horizontal projection correspond to background areas of the binary image. Furthermore, the amplitude and the frequency of the projection are maximized when the skew of the text is zero. Based on this characteristic, the image is rotated through a full range of angles. Finding peak in the horizontal projection gives the estimated angle of the text skew [2]. This circumstance is shown in Fig. 1.

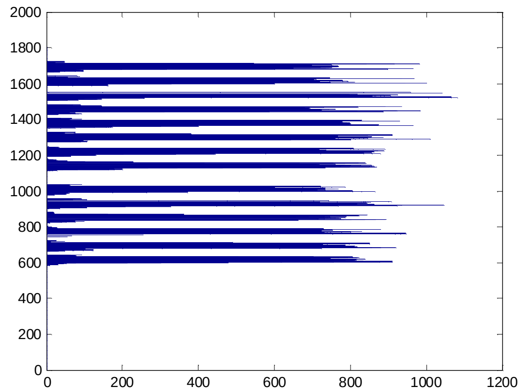
Moment based method

Moment defines the measure of the pixel distribution in the image. It identifies global image information concerning its contour. Moments of the binary image $B(i, j)$ are [3]:

$$m_{pq} = \sum_{i=1}^{N}\sum_{j=1}^{M} i^p j^q \,, \tag{2}$$



(a)



(b)

Fig. 1 Horizontal projection profiles method: (a) with skewed text, (b) with unskewed text

where p and q = 0, 1, 2, 3, ..., n, and n represent the order of the moment. From eq. (2) central moments μ_{pq} for binary image $B(i, j)$ can be calculated as:

$$m_{pq} = \sum_{i=1}^{N} \sum_{j=1}^{M} (i - \bar{x})^p (j - \bar{y})^q . \tag{3}$$

Some of the important image features are obtained from the moments. One of them is the object orientation defined by the angle θ. It represents the angle between the object and the horizontal axis. It can be estimated by minimizing the function [4]:

$$S(q) = \sum_i \sum_j [(i - \bar{x})\cos(q) - (j - \bar{y})\sin(q)]^2 . \tag{4}$$

The minimization gives the following values [3-4]:

$$q = \frac{1}{2}\arctan\left(\frac{2m_{11}}{m_{20} - m_{02}}\right). \tag{5}$$

All these features characterize the separate object. Hence, this method has been utilized regularly for the printed text skew evaluation. In that case, the skew evaluation of the each text object is similar or equal to all others.

Log-polar cross-correlation algorithm

The log-polar cross-correlation algorithm is based on the log-polar transformation. It emulates the principle of primate eye retina [5]. Each point in Cartesian space maps into equivalent point in the log-polar space. Firstly, the Cartesian space coordinates x and y are converted into polar coordinates radius r and angle θ. Their mapping is as follows [5]:

$$r = \sqrt{x^2 + y^2} , \tag{6}$$

$$q = \arctan\frac{y}{x} . \tag{7}$$

Furthermore, the log transformation is obtained as:

$$r = \ln r = \ln\sqrt{x^2 + y^2} . \tag{8}$$

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

The log-polar transformation is nonlinear and non-uniform. Nonlinearity has been introduced by polar mapping, while non-uniform sampling is the result of logarithmic scaling [6].

Let consider the digital form of the transformation. For the binary image **B**, the center point of transformation is given as $B(m_c, n_c)$. The radius is assigned as R. It ensures that the maximum number of pixels is included within the reference circle of the conversion. Accordingly, the center of the circle is given as [7]:

$$m_c = \frac{M}{2} , \quad n_c = \frac{N}{2} . \tag{9}$$

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) [6]:

$$r = \sqrt{(i - m_c)^2 + (j - n_c)^2} , \quad 0 \le r \le R , \tag{10}$$

$$q = \arctan\left(\frac{j - n_c}{i - m_c}\right), \quad 0° \le q \le 360° , \tag{11}$$

where $i = 1,\ldots,M$, $j = 1, \ldots, N$. The log-polar transformation is given as (ρ, θ), where ρ is given from eq. (8).

In the log-polar domain, the text image matrix **B** and referent object matrix **E** are marked as **BC** and **EC**. Cross-correlation shows a similarity measure between two images. It is [8]:

$$CC(q) = C_{coeff}\left(\mathbf{BC}, circshift(\mathbf{EC}, q)\right) , \tag{12}$$

where **ECS** is circshift(**EC**, θ) and C_{coeff} (**BC**, **ECS**) ia given as [8]:

$$C_{coeff} = \frac{\sum_r \sum_q (BC_{rq} - \overline{BC})(ECS_{rq} - \overline{ECS})}{\sqrt{\left(\sum_r \sum_q (BC_{rq} - \overline{BC})^2\right)\left(\sum_r \sum_q (ECS_{rq} - \overline{ECS})^2\right)}} . \quad (13)$$

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

The identification of the rotation in the image space is a complex task. In the log-polar space, the rotation 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 cross-correlated with the text image for the different angles, then it will be read out as the translation in the log-polar space. The first objective is the selection of the correct referent object as a template. The right choice can be the 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, the ellipse is split into left and right half part from the center point of the transformation. This way, the parts of the ellipse are matched with the text 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 the right one.

III. Experiments

The primary goal of the experiment is the evaluation of the algorithm used for the text skew estimation. Hence, it evaluates the algorithm's performance in the skew tracking domain. Experiments for the printed text consist mostly of the synthetic datasets. The most important one is the test of the single line printed text. This test consists of the single line printed text rotated for the angle θ from 0° to 60° in the steps of 5° around x-axis [9]. Typical text sample is shown in Fig. 2.

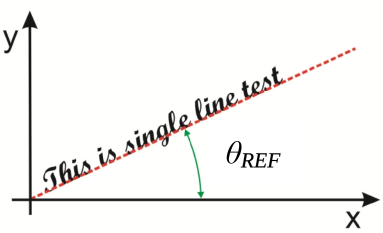


Fig. 2 Skew angle θ_A vs. referent angle θ_{REF}

For the evaluation of the method robustness, all text samples are given in the standard resolution of 300 dpi. The results are evaluated by the absolute, i.e. error. Absolute deviation is given as:

$$\Delta q_A = \left| q_{REF} - q_A \right| , \quad (14)$$

where θ_{REF} is the reference skew of the input text sample, while θ_A is the skew of the text sample obtained with tested algorithm.

IV. Results and Discussion

The test results for the different algorithms are given in Table I.

Table I
Absolute skew deviation for different methods

Angle range	Text skew estimation methods		
	Horizontal Projection method	Moment based method	Initial Skew rate method
θ_{REF} (°)	$\Delta\theta_{A,VPA}$	$\Delta\theta_{A,MA}$	$\Delta\theta_{A,LCCA}$
0	2.5	0.0000	0
5	0.1	0.0122	0
10	0	0.0089	0
15	0	0.0331	0
20	1.8	0.0435	0
25	0	0.0377	0
30	0	0.0417	0
35	0	0.0554	0
40	0	0.0248	0
45	0	0.0027	0
50	0	0.0175	1
55	0	0.0330	1
60	0	0.0545	1

The graph of the skew angle obtained by tested algorithms θ_A vs. reference angle θ_{REF} is shown in Fig. 3.

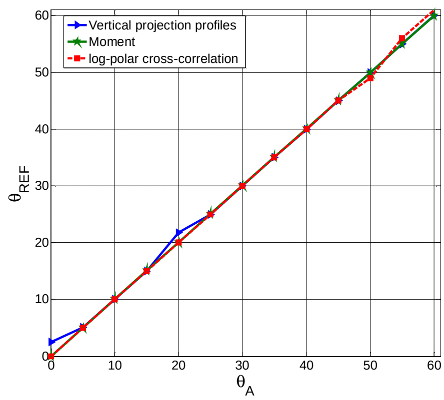


Fig. 3 Skew angle θ_A vs. referent angle θ_{REF}

Presented graph describes the response of the algorithm. If the algorithm has a correct response, then the graph will be linear. This means that the algorithm adequately identifies the text skew angle.

The vertical projection algorithm has the largest deviation (Fig. 3 - See blue line for reference). Log-polar cross-correlation algorithm has a linear response for the angles up to

45°. However, higher values of the angle θ_{REF} contributes to the nonlinearity (Fig. 3 - See red line for reference). In contrary, the moment based algorithm has uniform and linear response with a very small deviation in the whole range of angles. Hence, this method identifies the text skew angle with the smallest deviation. Fig. 4 shows the comparison of the absolute deviation results.

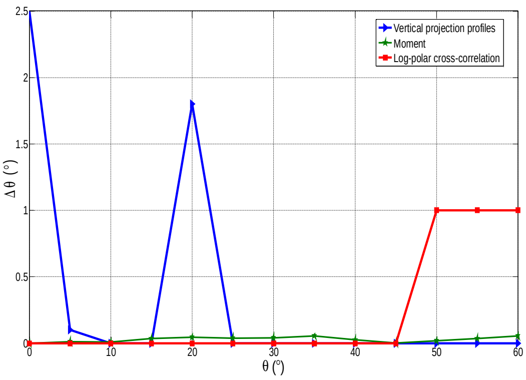


Fig. 4 Skew angle θ_A vs. reference angle θ_{REF}

The horizontal projection algorithm has the worst result. Its deviation is up to 2.5°. From the above, the log-polar cross-correlation algorithm has no deviation for the angles up to 45°. For the larger angle, it is up to 1°. Moment based algorithm has the uniform absolute deviation of up to 0.055°. This result is characterized as a pretty correct one. Summarized results for the tested algorithms are given in Table 2.

TABLE II
SUMMARY OF THE OBTAINED RESULTS FOR DIFFERENT METHODS

Angle range	**Text skew estimation methods**		
$0° \le \theta \le 60°$	**Horizontal Projection method**	**Moment based method**	**Initial Skew rate method**
$\Delta\beta$	< 2.5°	< 0.05°	< 1°

Taking into account the results from Tables 1-2, the moment based algorithm has the best result. It has a uniform absolute deviation and *RLHR*. Log-polar cross-correlation method is a promising one. However, it needs some adaptation. It is especially true for the angles higher than 45°. The horizontal projection algorithm has the worst results. Hence, it needs a serious adaptation.

However, tested algorithms have some additional obstacles in the application. Moment based algorithm has sensitivity to the noise. Hence, the additional cleaning of the image is the indispensable. Log-polar transformation is a very sensitive to the determination of the center point of transformation. If it is incorrect, then transformation will fail. Hence, it needs a special study on criteria for the center point determination. Vertical projection algorithm, in its initial form, has a problem with the correct text skew identification. In addition, its application is an easy one. Compared to the other methods, it is the slowest and error prone method. From above facts, the moment based algorithm is the best choice for the correct text skew estimation of the binary objects.

V. CONCLUSIONS

This paper gives the critical analysis, and comparison between different text skew estimation techniques, which include horizontal projection, moment based and log-polar cross-correlation method. All methods are tested with the document image samples in the standard resolution of 300 dpi. Analyzed algorithms show a different skew accuracy. The horizontal projection algorithm shows an obvious error prone. The log-polar cross-correlation algorithm has a satisfactory correctness. However, it needs some adaptation as well as a clear criteria for determination of transformation center point. Though it has obstacles, the moment based method has a clear advantage over the other methods. The most important is that it estimates correctly the text skew. Furthermore, it is computationally non-intensive method.

ACKNOWLEDGEMENT

This work is supported by a grant from the Ministry of Education, Science and Technological Development of the Republic of Serbia, as a part of the project No. TR33037.

REFERENCES

[1] A. Amin, S. Wu, "Robust skew detection in mixed text/graphics documents," Proceedings of 8th International Conference on Document Analysis and Recognition, pp.247–251, 2005.

[2] R. Manmatha, N. Srimal, "Scale space technique for word segmentation in handwritten manuscripts," Proceedings of 2nd International Conference on Scale Space Theories in Computer Vision, pp.22–33, 1999.

[3] P. Shivakumara, G. H. Kumar, H. S. Varsha, S. Rekha, M. R. Rashmi Nayaka, "A new moments based skew estimation technique using pixels in the word for binary document images," Proceedings of 8th International Conference on Document Analysis and Recognition, pp.151–156, 2005.

[4] G. Kapogiannopoulos, N. Kalouptsidis, "A fast high precision algorithm for the estimation of skew angle using moments," Proceedings of Signal Processing, Pattern Recognition, and Applications, pp.370–143, 2002.

[5] V. J. Traver, A. Bernardino, "A review of log-polar imaging for visual perception in robotics," Robotics and Autonomous Systems, vol.58, no.4, pp.378–398, 2010.

[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, vol.14, no.10, pp.1422–1434, 2005.

[7] M. K. Bhowmik, et all., "Classification of log-polar-visual eigenfaces using multilayer perceptron," Proceedings of 2nd International Conference on Soft computing, pp.12–23,2008.

[8] D. Brodić, Z. Milivojević, "Log-polar transformation as a tool for text skew estimation," Electronics and Electrical Engineering, vol.19, no.2, pp.61–64, 2013.

[9] D. Brodić, D. R. Milivojević, Z. N. Milivojević, "Basic test framework for the evaluation of text line segmentation and text parameter extraction," Sensors, vol.10, no.5, pp.5263–5279, 2010.