2013 12th International Conference on Document Analysis and Recognition

ICDAR2013 Document Image Skew Estimation Contest (DISEC'13)

A. Papandreou, B. Gatos, G. Louloudis and N. Stamatopoulos

Computational Intelligence Laboratory
Institute of Informatics and Telecommunications
National Center for Scientific Research "Demokritos"
GR-15310 Athens, Greece
{alexpap, bgat, louloud, nstam}@iit.demokritos.gr

Abstract **- The detection and correction of document skew is one of the most important document image analysis steps. The ICDAR2013 Document Image Skew Estimation Contest (DISEC'13) is the first contest which is dedicated to record recent advances in the field of skew estimation using well established evaluation performance measures on a variety of printed document images. The benchmarking dataset that is used contains 1550 images that were obtained from various sources such as newspapers, scientific books and dictionaries. The document images contain figures, tables, diagrams, architectural plans, electrical circuits and they are written in various languages such as English, Chinese and Greek. This paper describes the details of the contest including the evaluation measures used as well as the performance of the twelve methods submitted by ten different groups along with a short description of each method.**

Keywords – Skew estimation; document image preprocessing; performance evaluation; contest.

I. INTRODUCTION

In order to proceed with optical character recognition (OCR), document image skew correction is essential as a preprocessing step since some degree of skew is unavoidable to be introduced when a document is scanned manually or automatically [1]. The skew angle of a document image is defined as the deviation of the dominant orientation of the text lines from the horizontal axis. The existence of skew may seriously affect the performance of subsequent processing such as segmentation and OCR. Furthermore, a skew angle greater than 0.1° may be visible to a human observer. According to recent research, skew detection is still an interesting and challenging issue especially for documents with graphics, charts, figures or various font sizes [2]. In the literature, a variety of skew detection techniques are available and fall broadly into the following four categories according to the basic approach they adopt: projection profile based [3-7], Hough transform [8-12], nearest neighbor clustering [13-15] and interline cross correlation [16-20] methods. Several factors that restrict the efficiency of the above mentioned skew estimation methods are the unknown layout of the document image and the range of potential skew angles in which a method can efficiently estimate the skew.

In this first international skew estimation contest (DISEC'13) which is organized in conjunction with ICDAR2013, the general objectives are to make a comparison of current skew estimation techniques and to provide a dataset, missing from the literature, which could be considered as a generic benchmarking set.

For the construction of the datasets used in DISEC'13 (see Fig.1), we scanned, binarized [21] and rotated in a range of arbitrary angles 175 images from various types of documents, representative of most realistic cases that an algorithm might come up against. The document images used contain figures, tables, diagrams, block diagrams, architectural plans, electrical circuits, while they are obtained from newspapers, literature, comic, scientific and course books, dictionaries, travel and museum guides, official state documents and various other sources. The document images of the datasets are written mainly in English, Chinese and Greek languages, while there are several documents written in Japanese, Bulgarian, Russian, Danish, Italian, Turkish and ancient Greek languages. The sets contain representative cases of: (a) various sizes of document images, (b) any kind of mixed content, (c) vertical and horizontal writing, (d) multi-sized fonts and (e) multiple number of columns in the same document. All the 175 documents of the datasets were verified to have no skew and were randomly rotated in ten different angles, ranging from -15° to 15°, thus resulting in 1750 images with known ground-truth. At a next step, 200 representative samples of those images were selected to form the experimental dataset and were provided to the participants along with their ground-truth in order to tune their algorithms while the remaining 1550 document images formed the benchmarking dataset of the contest.

The contest procedure was based on the following milestones. The authors of candidate methods registered their interest in the contest and downloaded the experimental dataset. At a next step, all registered participants were required to submit their executables in the form of a console application. After the evaluation of all candidate methods, the benchmarking dataset (1550 document images along with

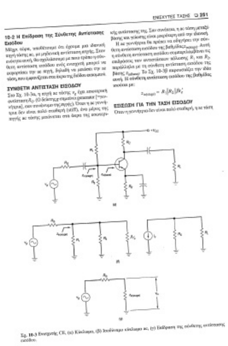
1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.291

the corresponding ground-truth information) became publicly available [22].

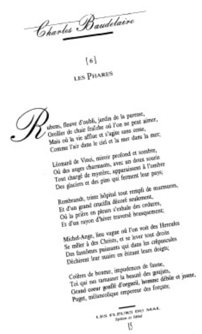
The remainder of the paper is organized as follows. In Section II, the participating groups together with a brief description of each method are summarized. Section III describes the performance evaluation protocol that was used while Section IV presents the experimental results of the contest. Finally, conclusions are drawn in Section V.

-3.92°

-0.07°

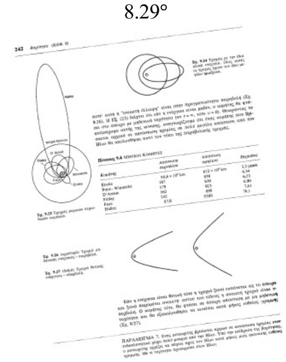


3.13°



8.29°

10.34°



7.44°

Figure 1. Image samples of the benchmarking dataset along with their corresponding skew angle.

II. METHODS AND PARTICIPANTS

Ten research groups submitted their methods to the contest. Two research groups submitted two different methods making the total number of participating methods equal to twelve. A brief description of these methods is provided in this section.

Ajou-SNU method: Submitted by Hyung Il Koo from Ajou University, Suwon, Korea and Nam Ik Cho from Seoul National University, Seoul, Korea and based on [27]. This method estimates the skew by detecting straight lines in gray-scale and binary document images. Therefore, it can take clues from text-lines, boundaries of figures, tables, vertical and horizontal separators as well as any combination of these entities. Specifically, a block-based edge detector that extracts several kinds of edges is developed. At a next step, straight lines are detected in edge maps and the skew angle is calculated by applying a maximum-likelihood estimation technique to the detected lines.

Aria method: Submitted by Farzad Nadi from the Department of Electrical and Computer Engineering, University of Hormozgan, Bandar-Abbas, Iran and Javad Sadri from the Department of Computer Engineering, University of Birjand, Birjand, Iran. This method uses the inversed version of the document image around the vertical and horizontal axis. The inversed image is put on the original image and the new combined result is rotated from -15° to 15°. At each rotation, the inversion of the rotated image around the vertical axis is obtained and put on the rotated image. Finally, the number of the common foreground pixels of the two images is counted and the skew angle of the document image is considered to be the angle that corresponds to the largest number of common foreground pixels.

CMC-MSU method: Submitted by Oleg Naydin from the Computational Mathematics and Cybernetics Lab of Lomonosov Moscow State University. This approach uses the Canny edge detector filter and Hough Transform in order to find lines on the document images. If there are no lines detected then it uses dilation and calculates the skew of each word.

CST-ECSU method: Submitted by Samit Biswas, Amit Kumar Das and Sekhar Mandal from the CST Department, Bengal Engineering and Science University, Shibpur, India and Bhabatosh Chanda, ECSU, Indian Statistical Institute, Kolkata, India. This method is based on Radon transform based projection profiles. At first, the document image is partitioned into several blocks. In Radon space, the maximum value or the highest peak determines the orientation of each block. A statistical distribution of skew angles of the blocks obtained through Radon transform is computed and the skew angle that corresponds to the mode of the distribution is taken as a rough estimation of the skew angle. At a next step, the document is deskewed accordingly and this procedure is repeated iteratively k-times to refine the estimated skew.

CVL-TUWIEN method: Submitted by Markus Diem, Florian Kleber and Robert Sablatnig from the Computer Vision Lab of Vienna University of Technology, Austria and Fraunhofer-Institute for Production Systems and Design Technology (IPK), Germany. This method is described in [23] and it is an extension of the Focus Nearest Neighbor Clustering (FNNC) proposed by Jiang et al. [24]. The method uses the Difference-of-Gaussians (DoG) interest points in order to detect a document's skew without the need of binarization, while the accuracy is increased by a voting based on straight lines and paragraph estimation.

Gamera method: Submitted by Christoph Dalitz from the Institute for Pattern Recognition, Niederrhein University of Applied Sciences, Germany and described in [26]. This algorithm is implemented for the Gamera framework for document analysis and recognition (gamera.sf.net), in the function rotation_angle_projections() and it is a variant of Postl's [3] projection profile method.

HIT-ICG-a method: Submitted by Xiangqian Wu, Youbao Tang and Hongyang Wang from the Image Computing Group, School of Computer Science and Technology, Harbin Institute of Technology (HIT-ICG), Harbin, China. This method is based on a line fitting technique for the four directions (up, down, left and right). For each direction, the original image is divided into 32 blocks. For each block, the minimum distance between the foreground pixels of the block and the corresponding direction boundary is computed. At a next step, these 32 points are used to conduct line fitting with roles. Finally, 4 fitting lines are detected. The skew angle of the best fitting line is considered as the skew angle of the original image.

HIT-ICG-b method: Submitted by the same group as the previous method. This method is based on the minimization of the cost of an energy function that takes under consideration the interactions between the bounding box and the outermost foreground pixels. The skew angle is obtained iteratively until the energy function has the minimum cost. In every iteration the bounding box is rotated decreasing the cost.

HP method: Submitted by Vandana Roy from Hewlett Packard, Bangalore, India. This method is a combined low complexity approach for skew detection using (a) paper edges in scanned image, (b) content boundaries based on Quasi-Hough Transform (QHT) and (c) scanned document image content based on e-PCP [19].

HS-Hannover method: Submitted by Karl–Heinz Steinke from the Hochschule Hannover, University of Applied Sciences and Arts, Germany and based on [25]. This approach detects horizontal and vertical lines for each angle between -15° and 15° in steps of 0.2°. Then, it determines the longest horizontal and vertical line around the best choice in steps of 0.01° in order to estimate the skew. If there are no horizontal and vertical lines contained in the image they are produced with the help of printed writing or handwriting.

LRDE-EPITA-a method: Submitted by Jonathan Fabrizio from the EPITA Research and Development Laboratory, Le Kremlin-Bicêtre, France. This method uses the magnitude spectrum of a frequency Fourier transform to determine the orientation of the document image. The document image is preprocessed and all regions of the document are clustered using a KNN. At a next step, the Fourier transform is applied on the image to all clusters convex hull boundaries. In that way, in the frequency domain, the orientation is easier to be detected.

LRDE-EPITA-b method: Submitted by Edwin Carlinet and Jonathan Fabrizio from the EPITA Research and Development Laboratory, Le Kremlin-Bicêtre, France and based on [28]. The submitted method detects contours of objects as a preprocessing step with the use of two filters. When the document image has enough structure, as line separators and frame borders, the Line Segment Detector (LSD) is used to detect the lines. In the case where the document has no structure, clustering of the connected components is taking place, giving lines or paragraphs from which the convex hulls are computed. The convex hulls of objects and lines from LSD are merged to give an image of meaningful segments. Finally, a standard Hough transform is applied on this document image to detect the skew angle.

III. PERFORMANCE EVALUATION

For every document image j of the benchmarking dataset the distance $E(j)$ between the ground-truth and the estimation of each submitted algorithm was calculated for each method. It should be noted that the estimations of each algorithm were rounded to the second decimal place as it was dictated by the contest's protocol. In order to measure the performance of the submitted methods the following three criteria were used: (a) the Average Error Deviation (AED), (b) the Average Error Deviation of the Top 80% of the results of each algorithm ($TOP80$) and (c) the percentage of Correct Estimations (CE). The definition of the above criteria is given in the rest of this section.

The AED criterion is described by:

$$AED = \frac{\sum_{j=1}^{N} E(j)}{N} \quad (1)$$

where N equals to 1550 and denotes the number of images of the benchmarking dataset.

For the calculation of the $TOP80$ criterion, the distances $E(j)$ were sorted, resulting in an ascending sE list, and the average error deviation is now calculated taking into account only the first 1240 values (80% of the images) of each list according to:

$$TOP80 = \frac{\sum_{j=1}^{M} sE(j)}{M} \quad (2)$$

where Mequals to 1240.

This criterion imprints the performance of each method excluding cases which we assume that the algorithm can't handle efficiently. In that way, the accuracy of the method is tested in its desired operation status. This is a criterion that was also used in [18, 20].

Finally, the CEcriterion is determined as:

$$CE = \frac{\sum_{j=1}^{N} K(j)}{N} \text{ where } K(j) = \begin{cases} 1 \ if \ E(j) \leq 0.1 \\ 0 \quad otherwise \end{cases}. \quad (3)$$

The threshold of 0.1° was chosen due to the fact that a skew angle greater than this threshold may be visible to a human observer.

For each criterion, the ranking of every submitted method was calculated. The final ranking is computed after sorting the accumulated ranking values for all criteria. Specifically, let $R(i)$ be the rank of the submitted method for the i^{th} criterion, where i =1,2,3. As denoted in Equation 4, for each skew estimation method, the final ranking S is achieved by the three rankings summation. The smaller the value of S the better performance is achieved by the corresponding method.

$$S = \sum_{j=1}^{3} R(j) \qquad (4)$$

IV. EVALUATION RESULTS

The performance of all participating algorithms was evaluated using the three criteria presented in the previous section. The evaluation results of all participating methods using the benchmarking dataset are presented in Table I while the ranking position of each method per criterion is presented in parentheses. The overall ranking of DISEC'13 participating methods, according to Eq.4, is presented in Table II and Figure 2.

TABLE I. EVALUATION RESULTS AND RANKING PER CRITERION

Method	*AED (°)*		*TOP80 (°)*		*CE (%)*	
Ajou-SNU	0.085	(2)	0.051	(2)	71.23	(2)
Aria	0.473	(8)	0.228	(12)	19.29	(12)
CMC-MSU	0.184	(5)	0.089	(10)	50.39	(10)
CST-ECSU	0.750	(10)	0.206	(11)	28.52	(11)
CVL-TUWIEN	0.103	(4)	0.058	(5)	65.42	(6)
Gamera	0.184	(5)	0.057	(4)	68.90	(3)
HIT-ICG-a	0.730	(9)	0.061	(6)	65.74	(5)
HIT-ICG-b	0.750	(10)	0.078	(9)	57.29	(9)
HP	0.768	(12)	0.073	(8)	58.32	(8)
HS-Hannover	0.227	(7)	0.069	(7)	58.84	(7)
LRDE-EPITA-a	**0.072**	(1)	**0.046**	(1)	**77.48**	(1)
LRDE-EPITA-b	0.097	(3)	0.053	(3)	68.32	(4)

TABLE II. OVERALL RANKING OF DISEC'13

Method	*AED*	*TOP80*	*CE*	*S*	*Overall Rank*
Ajou-SNU	2	2	2	6	2nd
Aria	8	12	12	32	11th
CMC-MSU	5	10	10	25	8th
CST-ECSU	10	11	11	32	11th
CVL-TUWIEN	4	5	6	15	5th
Gamera	5	4	3	12	4th
HIT-ICG-a	9	6	5	20	6th
HIT-ICG-b	10	9	9	28	9th
HP	12	8	8	28	9th
HS-Hannover	7	7	7	21	7th
LRDE-EPITA-a	1	1	1	3	1st
LRDE-EPITA-b	3	3	4	10	3rd

The best overall performance is achieved by **LRDE-EPITA-a** method which has been submitted by Jonathan Fabrizio from the EPITA Research and Development Laboratory, Le Kremlin-Bicêtre, France.

The ranking list for the first three methods is:

1. **LRDE-EPITA-a** (S = 3)
2. Ajou-SNU (S = 6)
3. LRDE-EPITA-b (S = 10)

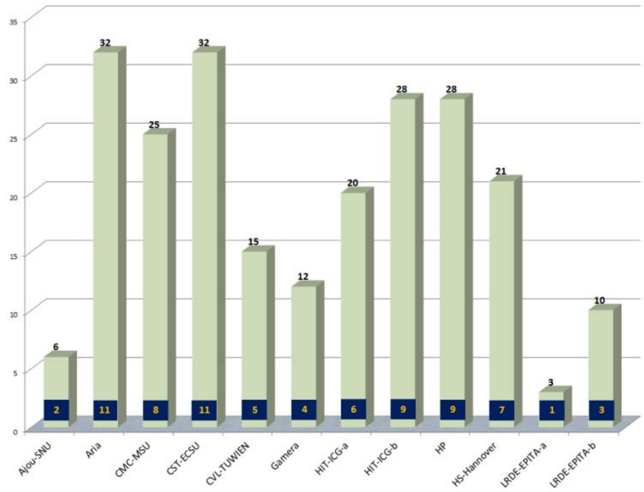


Figure 2. Overall ranking in terms of *S*.

After a careful analysis of the data presented in Tables I and II we can stress that:

a. The winning method (**LRDE-EPITA-a**) outperforms all other methods on all criteria that were used. Similarly, the second-ranked Ajou-SNU method is the second-best also using all three metrics.

b. Only three of the participating methods succeded to have an average error deviation under the well accepted threshold of 0.1°. On the other hand, most of the submitted algorithms (ten out of twelve) achieve this performance in the $TOP80$ criterion. This demonstrates that the participating techniques behave accurately in their desired operation status.

c. The comparison of the performance of each algorithm in AED and $TOP80$ criteria demonstrates whether there are numerous cases where each method majorly failed or the algorithm is robust and treats most of the cases in the same way. For example, in the comparison of AED and $TOP80$ criteria, the fact that **HP** and **HIT-ICG-a** methods gain four and three places respectively shows that they have in general better behavior than their average error but they majorly fail to handle certain cases. On the contrary, in the comparison of AED and $TOP80$ criteria, **CMC-MSU** and **Aria** methods lose five and four places respectively. This fact denotes that they haven't failed in specific cases rather they have a general high average error. Under this point of view, the remaining participating algorithms seem to be relatively robust since they tend to keep their rank in both criteria.

d. Although there are algorithms which operate with a relatively small average error deviation, they do not have similarly high performance in the CE criterion, where the highest performance achieved is 77.48%. All the algorithms have several cases that they can't adequately handle but there are more cases which are relatively easier and help them drop their average error.

Finally, for the top three ranked methods the standard error deviation around their average error was computed in order to measure their robustness. The results concerning **LRDE-EPITA-a**, **Ajou-SNU** and **LRDE-EPITA-b** are 0.06, 0.10 and 0.32 respectively and prove their robustness.

V. CONCLUSIONS

The ICDAR2013 Document Image Skew Estimation Contest (DISEC'13) is dedicated to record recent advances in the field of skew estimation in diverse documents using established evaluation measures. The benchmarking dataset of the contest was created using 155 representative document images that were obtained from various sources. In order to measure the accuracy of the submitted methods three criteria were used: (a) the Average Error Deviation, (b) the Average Error Deviation of the Top 80% of the results and (c) the percentage of Correct Estimations. Ten research groups participated in the contest with twelve submitted methods. The best overall performance is achieved by **LRDE-EPITA-a** method ($AED = 0.072°$, $TOP80 = 0.046°$ and $CE = 77.48\%$) which has been submitted by Jonathan Fabrizio from the EPITA Research and Development Laboratory, Le Kremlin-Bicêtre, France and uses the spectrum of a frequency Fourier transform to determine the orientation of the document image.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 600707 - tranScriptorium.

REFERENCES

[1] M. Sarfraz and Z. Rasheed, "Skew estimation and correction of text using bounding box", Proc. 5th International Conference on Computer Graphics, Imaging and Visualization, pp. 259–264, 2008.

[2] A. E. Sharif and N. Movahhedinia, "On skew estimation of Persian/ Arabic printed documents", Journal of Applied Sciences vol. 8 issue 12, pp. 2265–2271, 2008.

[3] W. Postl, "Detection of linear oblique structures and skew scan in digitized documents", Proc. 8th International Conference on Pattern Recognition, pp. 687-689, 1986.

[4] A. Papandreou and B. Gatos, "A Novel Skew Detection Technique Based on Vertical Projections", Proc. 11th International Conference on Document Analysis and Recognition, pp. 384-388, 2011.

[5] H. S. Baird, "The skew angle of printed documents", Proc. SPSE 40th Symposium Hybrid Imaging Systems, pp. 739–743M, 1987.

[6] G. Ciardiello, G. Scafuro, M. T. Degrandi, M. R. Spada and M. P. Roccotelli, "An experimental system for office document handling and text recognition", Proc. 9th International Conference on Pattern Recognition, pp. 739–743, 1988.

[7] Y. Ishitani, "Document skew detection based on local region complexity", Proc. 2nd International Conference on Document Analysis and Recognition, pp. 49–52, 1993.

[8] N. Srihari and V. Govindaraju, "Analysis of textual images using the Hough transfor", Machine Vision and Applications, pp. 141–153, 1989.

[9] J. Hinds, L. Fisher and D. P. D'Amato, "A document skew detection method using run-length encoding and the Hough transform", Proc. 10th International Conference Pattern Recognition, pp 464–468, 1990.

[10] J. Wang, M. K. H. Leung and S. C. Hui, "Cursive word reference line detection", Pattern Recognition vol. 30 issue 3, pp. 503–511, 1997.

[11] B. Yu and A. K. Jain, "A robust and fast skew detection algorithm for generic documents", Pattern Recognition vol. 29 issue 10, pp.1599–1629, 1996.

[12] C. Singh, N. Bhatia and A. Kaur, "Hough transform based fast skew detection and accurate skew correction methods", Pattern Recognition vol. 41, pp. 3528–3546, 2008.

[13] A. Hashizume, P. S. Yeh and A. Rosenfeld, "A method of detecting the orientation of aligned components", Pattern Recognition Letters, vol. 4, pp. 125-132, 1986.

[14] L. Gorman, "The document spectrum for page layout analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 15 issue 11, pp. 1162–1173, 1993.

[15] Y. Lu and C. L. Tan, "A nearest-neighbor chain based approach to skew estimation in document images", Pattern Recognition Letters vol. 24, pp. 2315–2323, 2003.

[16] H. Yan, "Skew Correction of Document Images Using Interline Cross-Correlation", CVGIP: Graphical Models and Image Processing, vol. 55, issue 6, pp. 538-543, 1993.

[17] B. Gatos, N. Papamarkos and C. Chamzas, "Skew detection and text line position determination in digitized documents", Pattern Recognition, vol. 30, issue 9, pp. 1505-1519, 1997.

[18] C. H. Chou, S. Y. Chu and F. Chang, "Estimation of skew angles for scanned documents based on piecewise covering by parallelograms", Pattern Recognition vol. 40, pp.443–455, 2007.

[19] P. Deya and S. Noushath, "e-PCP: A robust skew detection method for scanned document images", Pattern Recognition vol. 43, pp.937-948, 2010.

[20] A. Alireza, P. Umapada, P. Nagabhushan and F. Kimura, "A Painting Based Technique for Skew Estimation of Scanned Documents," Proc. 11th International Conference on Document Analysis and Recognition, pp. 299-303, 2011.

[21] B. Gatos, I. Pratikakis and S. J. Perantonis, "Adaptive Degraded Document Image Binarization", Pattern Recognition, vol. 39, pp. 317-327, 2006.

[22] http://www.iit.demokritos.gr/~alexpap/DISEC13/icdar2013_benchmarking_dataset.rar

[23] M. Diem, F. Kleber and R. Sablatnig, "Skew Estimation of Sparsely Inscribed Document Fragments", Proc. 10th IAPR International Workshop on Document Analysis Systems, pp.292-296, 2012.

[24] X. Jiang, H. Bunke, and D. Widmer-Kljajo, "Skew detection of document images by focused nearest-neighbor clustering", Proc. 5th International Conference on Document Analysis and Recognition, pp. 629 –632, 1999.

[25] K. H. Steinke, M. Gehrke and R. Dzido, "Recognition of Humboldt's Handwriting in Complex Surroundings", Proc. 12th International Conference on Frontiers in Handwriting Recognition, pp. 553-558, 2010.

[26] C. Dalitz, G. K. Michalakis and C. Pranzas, "Optical Recognition of Psaltic Byzantine Chant Notation", International Journal of Document Analysis and Recognition 11, pp. 143-158, 2008.

[27] H. I. Koo and N. I. Cho, "Skew estimation of natural images based on a salient line detector", Journal of Electronic Imaging, vol. 22, no. 1, 013020, 2013.

[28] R. Grompone von Gioi, J. Jakubowicz, J. M. Morel and G. Randall, "LSD: A Fast Line Segment Detector with a False Detection Control", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pp. 722-732, 2010.