

A robust skew detection method based on maximum gradient difference and R-signature

Mehdi Felhi
Océ Print Logic Technologies
LORIA UMR 7503 -
University of Nancy 2
Email: mehdi.felhi@oce.com

Nicolas Bonnier
Océ Print Logic Technologies
Email: nicolas.bonnier@oce.com

Salvatore Tabbone
LORIA UMR 7503 -
University of Nancy 2
Email: tabbone@loria.fr

Abstract—In this paper we study the detection of skewed text lines in scanned document images. The aim of our work is to develop a new automatic approach able to estimate precisely the skew angle of text in document images. Our new method is based on Maximum Gradient Difference (MGD) and R-signature. It detects zones that have high variations of gray values in different directions using the MGD transform. We consider these zones as being text regions. R-signature which is a shape descriptor based on Radon transform is then applied in order to approximate the skew angle.

The accuracy of the proposed algorithm is evaluated on an open dataset by comparing error rates.

Index Terms—Skew detection, text detection, document image analysis, maximum gradient difference, R-signature.

I. INTRODUCTION

Document image segmentation plays a crucial role in document image analysis as a pre-processing step. In fact, the performances of functional analysis depend mainly on the performances of the segmentation step. For instance, applying an Optical Character Recognition (OCR) [8] system directly on skewed document images without skew angle detection will result in a lot of failures in recognition rates. To improve results, several techniques have been proposed for the detection of skew. In this paragraph, we introduce some of these techniques. One of the most often used techniques is Projection Profiles (PP) (and extensions of PP) [1], [7], [10]. For instance, the method proposed by G. Nicchiotti et al. [10] introduces an extension of the PP called Generalized Projections (GP) able to distinguish among a set of black pixels which belong to noise from which belong to text. In this class of methods, the skew angle is estimated by minimizing a cost function applied in different angles. However, the performances of PP based methods depend directly on the type of the document image. For instance, S. Li et al. [7] have mentioned that these methods are very sensitive to the layout of the document image and are not robust to changes in font and size of text.

Other methods are based on connected components such as Nearest Neighbor based method (NN) [11] and Hough

Transform based methods (HT) [6], [11]. The NN methods have been used as a clustering method in order to estimate the skew angle. Concretely, a histogram of angles is calculated by determining the nearest neighbor of each connected component and the angle between them. Similarly, HT based methods draw the Hough plan of each document image and consider the peak value as the skew angle. These methods are also sensitive to the type of the document and are computationally expensive especially if the document contains a high number of connected components.

Furthermore, morphological mathematic based methods [9], [5] have been used for skew detection by applying morphological operators in order to transform text lines to a bounding box. But these methods are not robust to noise and to the choice of the parameters of morphological operators. Furthermore, most of the described skew detection approaches are dedicated to simple layout documents.

In this paper we introduce a new skew detection method able to extract text regions with different contrasts and independently from the layout and the type of document images. The result of this segmentation step is considered as an input to a Radon transform based shape descriptor which is R-signature in order to estimate the skew angle.

II. NEW APPROACH

A. Maximum Gradient Difference (MGD)

As we mentioned earlier, our method begins with a segmentation step in order to extract text regions. The segmentation process is based on the maximum gradient difference technique introduced in several video text extraction methods [15], [12].

In accordance with Wongs method [15], the first step involves calculating the horizontal gradient G of the image I :

$$G = I \star g, \quad (1)$$

where, $g = \begin{bmatrix} 1 & -1 \end{bmatrix}$.

This step is followed by selecting the maximum and the minimum values of the calculated gradient within a local window centered at each pixel p of size $w \times 1$.

$$\forall i, j, \text{MGD}(i, j) = \max(G(i, p), j - n < p < j + n) - \min(G(i, p), j - n < p < j + n),$$

where, $n = \frac{w}{2} - 1$.

Note that, even if the text is over a complex background or onto image regions, MGD values corresponding to text regions are often superior to MGD values of the background or images. The ideal window size depend directly on the characters size. In fact, authors of [15] claim that one of the best choices of the value w is a value that approximates the size of the characters in the text line. Specifically, the 1-D window should be slightly bigger than the characters length.

B. R-signature: A shape descriptor for skew angle detection

The R-signature is a shape descriptor that was introduced by Tabbone et al. [14] in 2005. This descriptor is based on Radon transform and represents a robust approach designed to identify complex shapes.

Let $I(x, y)$ be a image. The Radon transform [14] of the image I is defined by:

$$T_{R^I}(\rho, \theta) = \int \int I(x, y) \delta(x \cos(\theta) + y \sin(\theta) - \rho) dx dy, \quad (2)$$

where $\delta(\cdot)$ is defined as follows:

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

In our case we are interested only in the binary domain since we detect the text region in the preprocessing stage before the skew detection step.

Let D be a binary shape (a text region in our case). Then, the image I could be represented as follows:

$$f_D(x, y) = \begin{cases} 1 & \text{if } (x, y) \in D \\ 0 & \text{otherwise} \end{cases}$$

To be more explicit, Radon transform describes the scattering data obtained from length intersection of all the lines L_i with the function I for all θ_i and ρ (see Fig. 1).

Under these notations, R-signature is defined in [14] as follow:

$$R_I(\theta) = \int_{-\infty}^{+\infty} T_{R^I}^2(\rho, \theta), \quad (3)$$

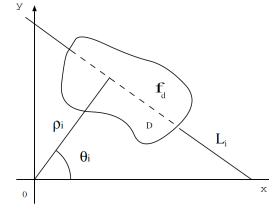


Fig. 1: Radon transform

C. Algorithm

The main idea of the proposed approach is to detect text lines in a compact representation and then estimate the skew angle (which is the orientation of the text lines) by means of R-signature. In order to improve precision rate of our method we adopted a successive refinement approach. We describe in this paragraph the algorithm of the proposed approach. Fig. 2 shows a sample result of the proposed algorithm.

- 1) At first, we calculate MGD_θ transform of the original image I in three directions $\theta \in \{\alpha_1, \alpha_2, \alpha_3\}$;

$$\text{MGD}_\theta = \text{Rotate}_{(-\theta)}(\text{MGD}(\text{Rotate}_\theta(I))), \quad (4)$$

where, $\text{Rotate}_\theta(I)$ is the rotation of the image I by an angle θ .

- 2) On each pixel location (i, j) we calculate the value minimizing MGD as follows :

$$\forall i, j, \text{MGD}_1(i, j) = \text{argmin}_\theta (\text{MGD}_\theta(i, j)), \quad (5)$$

The resulted MGD transform (MGD_1) is thresholded using a first threshold T_1 . Hence, we obtain a binary image B_1 (see Fig. 2b and Fig. 2c).

- 3) We estimate the skew angle θ_0 by means of R-signature: the first search size is set to 1° (see Fig. 2d);
- 4) We calculate then MGD in θ_0 direction and apply a second threshold $T_2 > T_1$ to calculate a second binary image B_2 (see Fig. 2e and Fig. 2f);
- 5) We search the final skew angle $\theta_f \in [\theta_0 - 1, \theta_0 + 1]$ by applying R-signature to B_2 : the final search size is set to 0.01° (see Fig. 2g).

D. Experimental results

We have evaluated our proposed method on the open dataset available on <http://ocrwks11.iis.sinica.edu.tw/~dar/Download/WebPages/Skew.htm> and provided by Chou et al. [4].

This dataset is composed of 500 document images generated by scanning a collection of different documents. These images are selected to represent different kind of documents (newspapers, books, magazines, and journals) and were produced by scanning different documents at 300 dpi. The designers of this database restricted the maximum possible angle to $\pm 15^\circ$.

Chou et al. [4] divided the dataset into 5 categories according to the language and the type of the documents:

- 1) English documents;

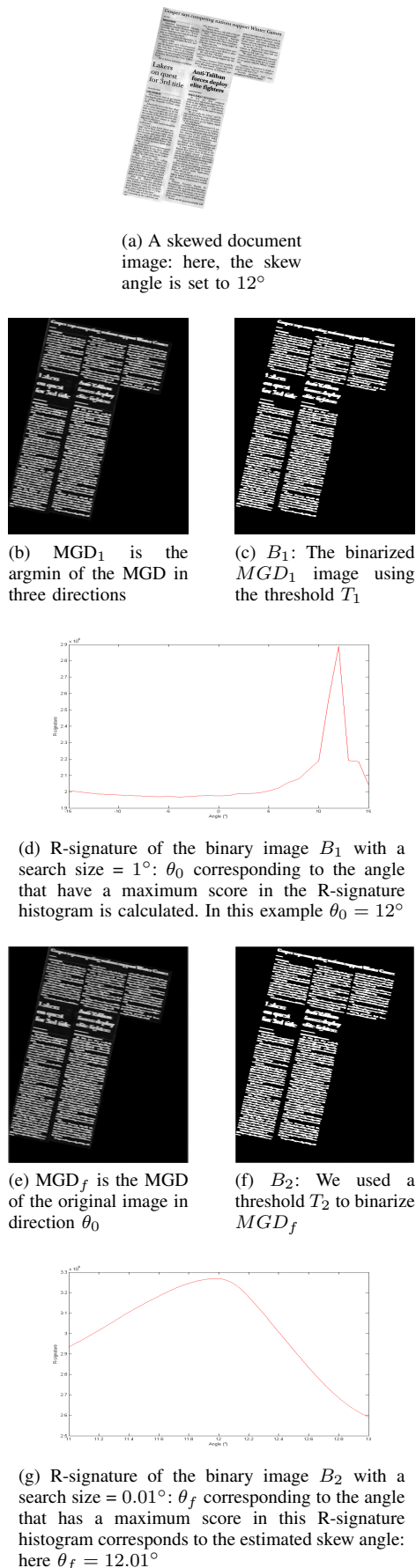


Fig. 2: Illustration of different steps of the proposed algorithm

- 2) Chinese and Japanese documents;
- 3) Documents containing large-scale figures;
- 4) Documents containing forms or tables;
- 5) Multilingual documents.

For comparison purpose, we selected five other existant methods; Wavelet [7], PCP [4], PJ [13], TC [3] and CC [2]. The experimental results of these methods on the dataset are available in [7] and [4].

In our evaluation, we decided to decrease the images resolution to 150 dpi in order to accelerate the calculation time. The parameter values of the algorithm are empirically determined: $w = 10$, $T_1 = 90$ and $T_2 = 100$. We also fixed $(\alpha_1, \alpha_2, \alpha_3) = (-10^\circ, 0^\circ, 10^\circ)$ to cover the interval $[-15^\circ, 15^\circ]$ well. The performance measures are the error mean and the error variance comparing to the ground truth established by the designers of the dataset. Tables I, II, III, IV and V show the performance of the five existing methods and the proposed method on each category of the database. These tables show also the performances of the top 80% error rates.

Comparing to the methods [7] and [4] that provide the best results in the literature for the adopted dataset provided by Chou et al. [4], we remark that our proposed method performs competitive performance rates especially in terms of variance error. In the next paragraph, we will discuss one of the advantages of the proposed method.

E. Multiple skew angles detection

In contrast to the methods [7] and [4], our approach is a local one that deals separately with the connected components (text lines provided by the MGD transform). We remarked that this property provides additional information comparing to [7] and [4]. In Fig. 3 we demonstrate that our approach is able to distinguish different dominant skew angles corresponding to the same image; we remark in Fig. 3c that the R-signature histogram presents two main peak values. These values approximate the orientation of each of the two skewed pages belonging to the original image of Fig.3a.

Method	Mean		Variance	
	All images	Top 80%	All images	Top 80%
Our approach	0.240	0.168	0.033	0.013
Wavelet	0.256	0.208	0.088	0.015
PCP	0.149	0.102	0.129	0.096
PJ	0.230	0.153	0.206	0.140
TC	0.185	0.148	0.180	0.131
CC	0.166	0.115	0.144	0.109

TABLE I: Performances on the 1st category

Method	Mean		Variance	
	All images	Top 80%	All images	Top 80%
Our approach	0.114	0.071	0.013	0.004
Wavelet	0.126	0.068	0.035	0.005
PCP	0.139	0.088	0.143	0.070
PJ	0.496	0.254	0.591	0.263
TC	0.171	0.108	0.155	0.091
CC	0.180	0.132	0.192	0.096

TABLE II: Performances on the 2nd category

Method	Mean		Variance	
	All images	Top 80%	All images	Top 80%
Our approach	0.488	0.440	0.022	0.014
Wavelet	0.499	0.450	0.019	0.011
PCP	0.231	0.178	0.135	0.011
PJ	7.787	3.419	9.049	4.934
TC	0.249	0.183	0.223	0.144
CC	0.345	0.223	0.325	0.186

TABLE III: Performances on the 3rd category

Method	Mean		Variance	
	All images	Top 80%	All images	Top 80%
Our approach	0.172	0.116	0.021	0.009
Wavelet	0.125	0.071	0.021	0.008
PCP	0.111	0.062	0.127	0.073
PJ	0.160	0.096	0.163	0.105
TC	0.150	0.078	0.180	0.084
CC	0.139	0.075	0.146	0.078

TABLE IV: Performances on the 4th category

Method	Mean		Variance	
	All images	Top 80%	All images	Top 80%
Our approach	0.154	0.083	0.037	0.004
Wavelet	0.071	0.040	0.006	0.002
PCP	0.077	0.051	0.075	0.050
PJ	2.050	0.208	5.816	0.264
TC	0.176	0.105	0.240	0.072
CC	0.197	0.129	0.230	0.125

TABLE V: Performances on the 5th category

III. CONCLUSION

We proposed a robust and precise skew detection method based on maximum gradient difference and R-signature. The maximum gradient difference serves to segment image into text regions and non-text regions, while the R-signature helps to estimate the skew angle. Experimental results show that our new skew detection method performs very well in terms of error variance which means that the proposed method is robust. Besides, we demonstrated that our method is able to detect two different dominant skew angles in the same document image. As a futur work, we plan to evaluate the performances of the proposed method to detect multiple dominant skew angles in scanned document images and to detect the skew in complex document images (that contain text regions over non-uniform backgrounds) since it begins with a segmentation step.

REFERENCES

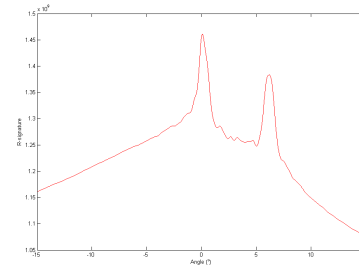
- [1] A. D. Bagdanov and J. Kanai. Projection profile based skew estimation algorithm for jbig compressed images. In *ICDAR*, pages 401–406. IEEE Computer Society, 1997.
- [2] A. Chaudhuri and S. Chaudhuri. Robust detection of skew in document images. *IEEE Transactions on Image Processing*, 6(2):344–349, 1997.
- [3] Y. K. Chen and J. F. Wang. Skew detection and reconstruction based on maximization of variance of transition-counts. *Pattern Recognition*, 33(2):195–208, February 2000.
- [4] C.-H. Chou, S.-Y. Chu, and F. Chang. Estimation of skew angles for scanned documents based on piecewise covering by parallelograms. *Pattern Recognition*, 40(2):443–455, 2007.
- [5] A. K. Das and B. Chanda. A fast algorithm for skew detection of document images using morphology. *IJDAR*, 4(2):109–114, 2001.
- [6] D. S. Le, G. R. Thoma, and H. Wechsler. Automated page orientation and skew angle detection for binary document images. *Pattern Recognition*, 27(10):1325–1344, 1994.
- [7] S. Li, Q. Shen, and J. Sun. Skew detection using wavelet decomposition and projection profile analysis. *Pattern Recognition Letters*, 28(5):555–562, 2007.
- [8] S. Mori, C. Y. Suen, and K. Yamamoto. Historical review of ocr research and development. *Proceedings of the IEEE*, 80(7):1029–1058, jul 1992.
- [9] L. Najman. Using mathematical morphology for document skew estimation. In Elisa H. Barney Smith, Jianying Hu, and James Allan, editors, *DRR*, volume 5296 of *SPIE Proceedings*, pages 182–191. SPIE, 2004.
- [10] G. Nicchiotti and C. Scagliola. Generalized projections: A tool for cursive handwriting normalization. *Document Analysis and Recognition, International Conference on*, 0:729, 1999.
- [11] L. O’Gorman. The document spectrum for page layout analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11):1162–1173, 1993.
- [12] T. Q. Phan, P. Shivakumara, and C. L. Tan. A laplacian method for video text detection. In *ICDAR*, pages 66–70. IEEE Computer Society, 2009.
- [13] W. Postl. Detection of linear oblique structures and skew scan in digitized documents. In *ICPR*, pages 687–689, 1986.
- [14] S. Tabbone, L. Wendling, and J.-P. Salmon. A new shape descriptor defined on the radon transform. *Computer Vision and Image Understanding*, 102(1):42–51, 2006.
- [15] E. K. Wong and M. Chen. A new robust algorithm for video text extraction. *Pattern Recognition*, 36(6):1397–1406, 2003.



(a) An example of a document image I that presents two dominant skew angles: the first skew angle is equal to 6° while the second is set to 0°



(b) $MGD_1(I)$



(c) R-signature of the binary image B calculated by thresholding $MGD_1(I)$: Two main peaks corresponding to the angles 6° and 0° are detected

Fig. 3: Localizing two different dominant skew angles