# Fast and Accurate Skew Estimation Based on Distance Transform

Itay Bar-Yosef, Nate Hagbi, Klara Kedem
Computer Science Department
Ben-Gurion University
Beer-Sheva, Israel
{itaybar, natios, klara}@cs.bgu.ac.il

Itshak Dinstein
Computer and Electrical Engineering
Ben-Gurion University
Beer-Sheva, Israel
dinstein@ee.bgu.ac.il

## Abstract

*Document skew estimation is an important step in the process of document analysis. In this paper we discuss the properties of the distance transform of binarized documents and derive a fast and accurate method for detecting document skew. The method is based on the observation that the dominant orientation of the gradient of the distance transform accurately reflects the skew of the document. Our experiments suggest the method is robust to large variations in text properties, such as difference in script and page layout, and provides estimation accuracy of state-of-the-art skew detection methods.*

## 1. Introduction

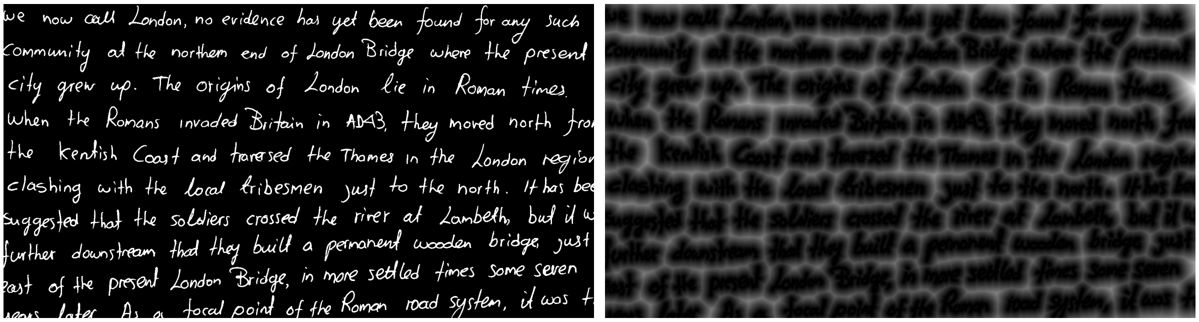
Skew deviations often occur in the document capturing process and affect the performance of subsequent stages, such as line extraction, page segmentation, and OCR. Although skew estimation has been widely researched throughout the years, it is still an active research area.

Skew estimation methods can be divided into several main categories. These include the popular projection profile methods [1, 2, 3], Hough transform based methods [4, 5], and nearest-neighbor based methods [6, 7, 8]. For a comprehensive survey of skew detection methods refer to [9]. Projection profile based methods are appropriate for small skew angle ranges, since computation time directly depends on the range and resolution of the skew angle. Methods based on Hough transform give accurate results, but require high computation time and large memory space. Nearest-neighbor based methods perform connected-component analysis, which is highly time consuming. In addition, such methods require special attention for dealing with different scripts, and connected or broken characters, and heavily depend on the quality of the binarization process output. This dependency can be very problematic when dealing with complex or degraded data, such as historical documents.

Gradient based methods [10, 11], also addressed in Cattoni's survey [9], received relatively little attention in the literature. Sun *et al.* [10] use the histogram of image gradient directions to find the dominant orientation in the document image. After histogram smoothing, the peak of the histogram indicates an estimate of the skew angle. Sauvola *et al.* [11] propose a similar approach in which directional Gaussian kernels are convolved with the image. This provides an approximation of gradient direction. The histogram of these directions is then used to find the skew angle. The main assumption of these methods is that the dominant gradient orientation of the foreground is perpendicular to text lines. Although this assumption is generally correct, using it directly may meet different accuracy problems, e.g. in the cases of degraded text and cursive handwriting. More recently, several interesting works have been reported by the group of Tan *et al.* [12, 13]. Yuan and Tan [12] propose an approach based on the connected-component grouping of convex hulls of different document components. After the grouping process, a histogram of the convex hull edge slopes is calculated and the highest peak of the histogram corresponds to the skew angle. Lu and Tan [13] propose an algorithm based on the analysis of horizontal and vertical white run length histograms of the background. They base their approach on the observation that text images normally hold a large amount of equidistant interline spacings. The skew is estimated using the white runs that exactly span the interline spacing.

In this paper we propose a method for skew estimation based on analyzing the distance transform (*DT*) of the document's foreground. We show that the dominant orientation of the *DT*'s gradients provides a highly accurate estimate of the skew angle. The proposed approach is simple, elegant, and highly robust to large variations in text properties, such as script and page layout. The method can be efficiently implemented, since it does not require any time consuming calculations, such as connected component labelling, projection to multiple angles, or Hough transform.

978-0-7695-3337-7/08 $25.00 © 2008 IEEE
DOI 10.1109/DAS.2008.65




(a) (b)

**Figure 1. (a) Document image (b) Corresponding distance transform image.**

The calculation of the *DT* can be implemented very efficiently, especially using dedicated hardware, such as *GPU* (Graphics Processing Unit). For a recent survey on 2D Euclidean distance transform and fast implementations, refer to [14].

This paper is organized as follows. Section 2 describes the properties of the distance transform of text documents and the details of our algorithm. Section 3 describes experimental results and Section 4 concludes our work and outlines future research directions.

## 2. Proposed Approach

In this paper we propose a fast and accurate method for calculating document skew. The proposed method is based on the distance transform and the observation that the dominant orientation of its gradients accurately reflects the skew of the document.

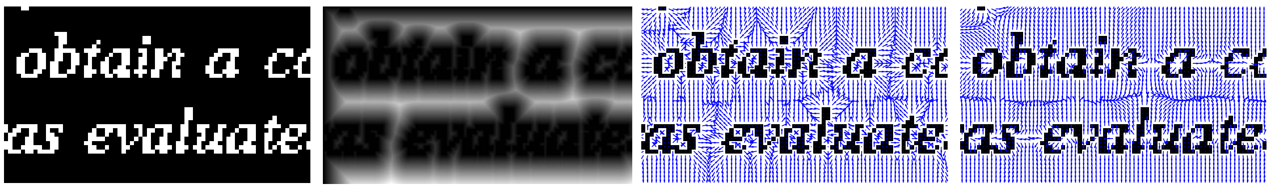
The proposed approach can be considered as a background analysis method since it involves analyzing the properties of the distance transform, which is propagated from the document foreground to the background. Background analysis has been extensively used in the literature and has been proven to be highly effective for different document analysis tasks, such as page segmentation [15, 16]. However, very few works have used background analysis to calculate document skew [13, 17]. There are several advantages for analyzing the document's background, rather than solely the text itself. Background analysis methods are less sensitive to text degradation, and are generally independent of text properties, such as script, font type and size, and page layout. In addition, connected characters or merged text lines highly affect skew methods which are based on text analysis. On the other hand, background methods can exploit information derived from the space surrounding the text, thus overcoming this problem.

The *DT* of a binary image is an image in which the value of each background pixel represents its Euclidean distance to the nearest foreground pixel. Figure 1 shows a document image and its corresponding *DT* image. In the following section we describe the important properties of the distance transform of a text document and their relation to the skew of the document.

### 2.1 Distance Transform Orientation

The *DT* of a binary document contains much important information about the structure of the document. As can be seen in Figure 1(b), the *DT* image has local maxima between neighboring lines. The *DT* image values increase from the boundaries of the foreground towards local maxima. The gradient vectors of the *DT* image point to the nearest local maxima, i.e. towards the central line of gaps. This is illustrated in Figure 2, which shows a portion of a document image, its *DT*, and the gradient vectors of the *DT*.

Considering the orientation of gradient vectors in the *DT* image, disregarding their sign (direction), the dominant orientation is perpendicular to the orientation of the text lines. Following this motivation, our method is aimed at estimating the dominant orientation of the gradient vectors of the *DT* image. Notice that gradient vectors in gaps between characters in the same text line converge to the local maximum between the characters (for example, notice the gap between the word "obtain" and the letter "a" in Figure 2(c)). These gradient vectors can potentially form a different dominant orientation. This effect can be reduced by convolving the *DT* with a Gaussian filter. The smoothing reduces the local maxima inside lines, and increases the convergence in gaps between them (see Figure 2(d)).



(a) (b) (c) (d)

**Figure 2. (a) A portion of a text document image (b) The $DT$ of the document image (c) Gradient orientation field of the $DT$ (d) Gradient orientation field of the smoothed $DT$. Notice the convergence of the gradient vectors towards the central line of the gap between neighboring text lines in the orientation field of the smoothed $DT$.**

### 2.2 Skew Angle Estimation

In this section we describe the algorithm in detail. Given a gray scale document image, we first binarize it using Otsu's global thresholding approach [19].

Next, we calculate the *DT* image of the binarized document image using the linear time algorithm proposed by Heinz et al [18]. As explained in Section 2.1, we convolve the *DT* image with a Gaussian filter, resulting in a smoothed *DT* image, denoted by $d_s$. The next step is calculating the gradient of $d_s$ in each background pixel

$$\nabla d_s = (\nabla_x, \nabla_y) = (\frac{\partial d_s}{\partial x}, \frac{\partial d_s}{\partial y}). \tag{1}$$

The gradient direction of $d_s$ can be approximated by

$$\theta = tan^{-1}(\frac{\nabla_y}{\nabla_x}), \theta \in [-90, 90]. \tag{2}$$

To robustly estimate the orientation in each pixel, most methods divide the image into equal-sized $N$x$N$ windows and average the orientation in each window (in all our experiments we used $N = 12$). However, since the dominant orientation gradient vectors between text lines converge to the center of the gap from two opposite directions, they are expected to cancel each other. To solve this problem, Kass and Witkin [20] propose to double the angles before the averaging process, so that ($\theta$+180) turns into (2$\theta$+360), which is equal to 2$\theta$. This allows us to effectively consider the orientation of the gradient vectors, disregarding their direction. In order to calculate gradient orientation, we adopt the approach proposed by Bazen *et al.* [21], which showed that the main orientation of an $N$x$N$ block can be calculated by

$$\phi = \frac{1}{2}\angle(\nabla_x^2 - \nabla_y^2, 2\nabla_x\nabla_y) \tag{3}$$

where $\angle$ is defined as:

$$\angle(x,y) = \begin{cases} tan^{-1}(\frac{y}{x}) & , \quad if\ x \geq 0 \\ tan^{-1}(\frac{y}{x}) + \pi & , \quad if\ x < 0 \land y \geq 0 \\ tan^{-1}(\frac{y}{x}) - \pi & , \quad if\ x < 0 \land y < 0 \end{cases} \tag{4}$$

and the average orientation $\theta$, which is perpendicular to $\phi$, is

$$\theta = \begin{cases} \phi + \frac{1}{2}\pi & , \quad for\ \phi \leq 0 \\ \phi - \frac{1}{2}\pi & , \quad for\ \phi > 0 \end{cases} \tag{5}$$

This gives an estimation for the orientation in the range $[-90°, 90°]$.

As mentioned earlier, our method is based on the observation that the dominant orientation of the *DT* gradient vectors is perpendicular to text lines. In order to estimate the dominant orientation, we thus calculate a histogram, $h_\theta$, for the orientations obtained by Equation (5). The angle that corresponds to the peak of $h_\theta$ is the estimated skew angle $\theta_s$:

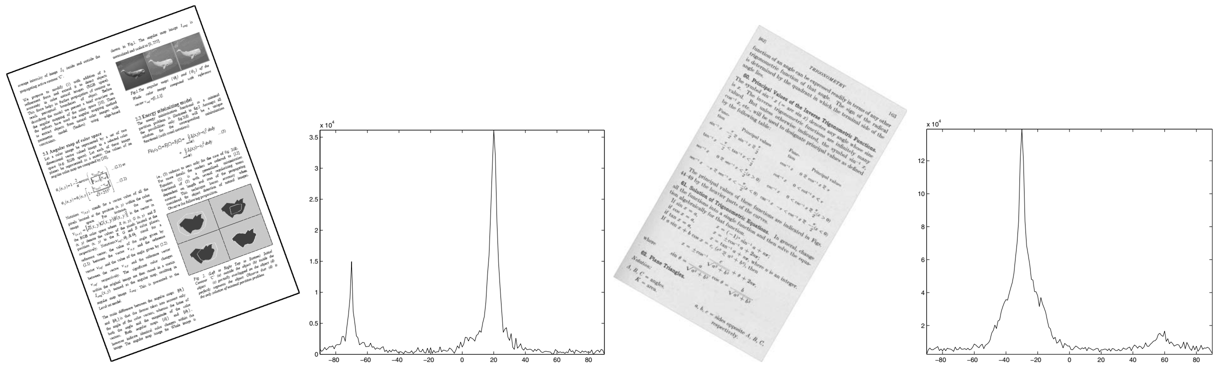
$$\theta_s = argmax_\theta(h_\theta) \tag{6}$$

The number of bins used to calculate $h_\theta$ determines the accuracy to which the dominant orientation is calculated. In our experiments we used $18000$ bins to represent $\theta$ in the range $[-90, 90]$, which provides a resolution of up to $0.01°$ for $\theta$. To reliably estimate the peak of $h_\theta$ we fit a Gaussian over $h_\theta$ and take its central value.

Figures 3(a) and 3(c) show two document images rotated at $20°$ and $-30°$ respectively. Figures 3(b) and 3(d) show the corresponding histograms $h_\theta$. The clear peak in $h_\theta$ corresponds to the dominant orientation, which is the desired skew angles.

## 3. Experimental Results

We conducted our experiments on three sets of documents (see Figure 4). The first set contains $40$ document



(a) (b) (c) (d)

**Figure 3. (a) A document image rotated at $20^\circ$ (b) Corresponding histogram $h_\theta$. (c) A document image rotated at $-30^\circ$ (d) Corresponding histogram $h_\theta$. Notice the prominent peaks of the histograms, which accurately corresponds to the true skew angles.**

images extracted from scientific articles, containing text, tables, and graphics. The second set contains 20 Chinese document images, extracted from scientific journals, and the last set contains 10 Hebrew handwritten documents, containing cursive text and several hand drawn figures. The Hebrew cursive documents are written on a checker notebook. The checker's parallel lines were used to establish the ground truth angle for the cursive documents, and were removed before the skew estimation process.

In order to evaluate the performance of our skew estimation algorithm, the 70 documents were rotated by 30 different skew angles, ranging from -90° to 90°, providing $70 \times 30 = 2100$ test images. We compared the results of the proposed approach with the projection profile method of Postl *et al.* [3], and with the gradient based approach of Sauvola *et al.* [11]. The projection profile method of Postl was implemented in an hierarchical manner, i.e., starting from a search resolution of $5^\circ$, then $1^\circ$, $0.5^\circ$, $0.1^\circ$ and finally resolution of $0.01^\circ$. The different methods were implemented in MATLAB.

The projection profile method gives highly accurate results for documents that are well structured and consist of mostly text. However, for documents of complex layout and different types of figures, such as the document in Figure 4(a), the projection profile method gives arbitrary large errors. Sauvola's method yielded acceptable results (less than $1^\circ$ error in average). However, the presence of graphics and different fonts decreases its accuracy. Our approach operated accurately and robustly, regardless of the document script and layout.

The results of our experiments are summarized in Table 1. The mean and standard deviation of the skew estimation errors are reported for our method, Sauvola's [11], and Postl's [3].

| Data set | Method | Sauvola | Postl | Ours |
|---|---|---|---|---|
| English | Mean error | 0.64 | 0.11 | **0.039** |
| | Std error | 0.25 | 0.21 | **0.026** |
| Chinese | Mean error | 0.82 | 0.26 | **0.035** |
| | Std error | 0.36 | 0.29 | **0.039** |
| Cursive | Mean error | 0.50 | 0.29 | **0.055** |
| | Std error | 0.21 | 0.16 | **0.042** |

**Table 1. Mean and standard deviation of the skew estimation errors. Our method gives highly accurate results, regardless of the documents class.**

# 4. Conclusion and Future Work

In this paper we proposed a fast and accurate gradient based method for document skew estimation. Our method is based on the distance transform (*DT*) of the binarized document, which contains important information about the document. The main observation behind the method is that the dominant orientation of the gradient vectors of the *DT* accurately reflects the skew of the document. We described an algorithm for extracting the dominant orientation of the *DT* in a robust and accurate manner. Our experimental results show the method is robust to variations in text properties, such as difference in script and page layout, and provides state-of-the-art accuracy. In addition, the proposed method

(a) (b) (c)

**Figure 4. Examples from the three sets of documents used in our experiments (a) A document image containing both text and graphics extracted from a scientific journal (b) A Chinese document image (c) A cursive Hebrew handwritten document.**

can be implemented very efficiently using dedicated hardware, such as *GPU*. We plan to perform a thorough comparative evaluation with different methods on larger data sets.

As pointed out by Antonacopoulos [17], analyzing the background gives rise to several issues. The lack of background space in text documents, such as documents containing dense and touching text lines, can lead to wrong skew estimation. However, this also affects methods which are based on text analysis. In our approach, even a small background space surrounding the dense text region, will suffice for correct skew estimation.

Another issue is the affect of oriented figures. Since our method is based on analyzing the orientation of the background, skew estimation of documents containing only a few text lines, will be affected by oriented figures. We note that this limitation is common to most skew detection techniques.

The *DT* of a document contains much information about the document. We intend to investigate methods for extracting and manipulating this information for other document analysis tasks, such as page segmentation and curved line segmentation.

## 5. Acknowledgments

This work was partially supported by the Lynn and William Frankel Center for Computer Sciences, the Israel Ministry of Science and Technology, and by the Paul Ivanier Center for Robotics and Production Management at Ben-Gurion University, Israel.

## References

[1] H. S. Baird. “The skew angle of printed documents”. Proceedings of Society of Photograhic and Engineers, pp. 14-21, 1987

[2] D. Bloomberg, G. Kopec, and L. Dasari. “Measuring document image skew and orientation”. SPIE 2422, pp. 302-316, 1995

[3] W. Postl, “Detection of linear oblique structures and skew scan in digitized documents”, ICPR 1986, pp. 687–689.

[4] A. Amin and S. Fischer, “A Document skew detection method using the Hough transform”. Pattern Analysis and Applications, Vol 3 , pp. 243-253, 2000

[5] S.N. Srihari, V. Govindaraju, “Analysis of textual images using the Hough transform”, Mach. Vision Appl. Vol 2, 1989, pp. 141-153

[6] L. OGorman, “The document spectrum for page layout analysis”, IEEE Trans. Pattern Anal. Mach. Intell. Vol 15, 1993, pp. 1162-1173

[7] 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.

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

[9] R. Cattoni, T. Coianiz, S. Messelodi, C.M. Modena, "Geometric Layout Analysis Techniques for Document Image Understanding: a Review", ITC-IRST Technical Report No. 9703-09, 1998.

[10] C. Sun and D. Si. "Skew and Slant Correction for Document Images Us- ing Gradient Direction". In Proc. of the 4th International Conference on Document Analysis and Recognition, pp. 142–146, Ulm, Germany, August 1997.

[11] J. Sauvola and M. PietikAainen. "Skew Angle Detection Using Texture Direction Analysis". In Proc. of the 9th Scandinavian Conference on Image Analysis, pp. 1099–1106, Uppsala, Sweden, June 1995.

[12] B. Yuan, C. L. Tan, "Convex hull based skew estimation", Pattern Recognition, Vol 40, pp. 456–475, 2007.

[13] S. J. Lu, J. Wang, C. L. Tan, "Fast and Accurate Detection of Document Skew and Orientation", ICDAR 2007, Vol 2, pp. 684–688.

[14] R. Fabbri, L. Costa, J. Torelli, O. Bruno, "2D Euclidean distance transform algorithms: A comparative survey", ACM Computing Surveys, Vol 40, pp. 1–44, 2008.

[15] K. Kise, O. Yanagida, S. Takamatsu, "Page segmentation based on thinning of background", ICPR 13, Vol 3, pp. 788–792, 1996.

[16] H.S. Baird, "Background structure in document images". Document Image Analysis, World Scientific, 1994, pp. 17-34

[17] A. Antonacopoulos, "Local skew angle estimation from background space in text regions", ICDAR 1997, Vol 2, pp. 684–688.

[18] B. Heinz, G. Joseph, D. Kirkpatrick, and M. Werman, "Linear Time Euclidean Distance Transform Algorithms,". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 5, May 1995, pp. 529–533

[19] N. Otsu, "A threshold selection method from gray-level histogra", IEEE Transaction on System Man and Cybernetics, Vol. 9, pp. 62–66, 1979.

[20] M. Kass, A. Witkin, "Analyzing oriented patterns", Comput. Vision, Graph. Image Process. Vol. 37 pp. 362-385, 1987

[21] A.M. Bazen, S.H. Gerez, "Systematic methods for the computation of the directional fields and singular points of fingerprints", IEEE Trans. Pattern Anal. Machine Intell. Vol. 24 pp. 905-919, 2002