

Estimation of Skew Angle in Binary Document Images Using Hough Transform

Nandini N., Srikanta Murthy K., and G. Hemantha Kumar

Abstract—This paper includes two novel techniques for skew estimation of binary document images. These algorithms are based on connected component analysis and Hough transform. Both these methods focus on reducing the amount of input data provided to Hough transform. In the first method, referred as word centroid approach, the centroids of selected words are used for skew detection. In the second method, referred as dilate & thin approach, the selected characters are blocked and dilated to get word blocks and later thinning is applied. The final image fed to Hough transform has the thinned coordinates of word blocks in the image. The methods have been successful in reducing the computational complexity of Hough transform based skew estimation algorithms. Promising experimental results are also provided to prove the effectiveness of the proposed methods.

Keywords—Dilation, Document processing, Hough transform, Optical Character Recognition, Skew estimation, and Thinning.

I. INTRODUCTION

OPTICAL character recognition belongs to the most important image analysis tasks. Its main applications are in building digital libraries (containing text, math formulas, music scores etc.), recognizing objects on digitalized maps, localization of vehicle license plates, text readers for visually impaired people, understanding hand-written office forms, bank checks etc. A typical OCR system consists of the following steps:

- image preprocessing, e.g. noise attenuation, correction of image orientation;
- image binarization, usually performed adaptively;
- segmentation, usually hierarchical (recognizing page layout, detecting text areas (and tables, figures etc.), then text paragraphs, individual lines, then segmenting lines into words, and finally words into characters);
- actual recognition (supervised or unsupervised);
- spellchecker-guided postprocessing;
- saving output in some popular format (html, PDF, LaTeX).

Preprocessing is a stage in typical OCR system, which focuses on enhancing the acquired image to increase the ease of

feature extraction and to compensate for the eventual poor quality of the scanned document. The subsequent stages of OCR systems mainly depend upon the accuracy of preprocessing stage. The mainstream commercial OCR tools are optimized for scanner-captured rather than camera-captured documents. During the scanning process, the whole document or a portion of it is fed through a scanner. The digital image of a document may be skewed/rotated arbitrarily because of how it was placed on the platen when it was scanned or because of a document feeder malfunction. A significant skew in document can be detected by human vision easily and the skew correction can be made by re-scanning the document, whereas for mild skew it may not be possible to notice its skew as human vision system fails to identify it. Even a smallest skew angle existing in a given document image results in the failure of segmentation of complete characters from words or a text lines, as the distance between the character reduces. Further, most of the OCRs and document retrieval/display systems are very sensitive to skew in document images. Hence it is important to detect and correct skew.

Intensive research work in the field of skew detection has given birth to many methods. The majority of them are based on Projection profile, Fourier transform, Cross-correlation, Hough transform, Nearest neighbor connectivity, Linear regression analysis and Mathematical morphology. A method for skew detection, using projection profile is proposed by [1], in which number of projections are obtained at different angles close to the best expected orientation and variations are observed for each projection. The maximum peak in the projection with best match to the text lines is considered to be skew angle. This method fails for documents with multi-column layout. Further, the accuracy reduces if the document contains noise.

In [2], a method based on Fourier Transform is proposed. In this method the angle of direction in which the density of Fourier space is large is considered to be the skew angle. The time complexity for this method is considerably large especially when the images are of bigger sizes.

In [3], a skew detection method using the cross correlation between the text lines at a fixed distance which is based on the fact that the correlation between vertical lines in an image is maximum for a skewed document, is presented. It is found that the proposed method is computationally expensive and gives lesser accuracy.

A bottom up technique for skew estimation based on nearest neighbor clustering is proposed by [4]. In this method, nearest neighbors of all the connected components are determined. The direction vectors for all the pairs of the nearest neighbors are accumulated in a histogram. The histogram peak is found to

Manuscript received June 30, 2008.

Nandini N. is with the Department of Computer Science & Engineering, PES Institute of Technology, Bangalore, India.

Srikanta Murthy K is with the Department of Information Science & Engineering, PES Institute of Technology, Bangalore, India.

G Hemantha Kumar is with the Department of Studies in Computer Science, University of Mysore, Mysore, India.

give skew angle. Since only one nearest neighbor connectivity is made for each component, connection with noisy sub parts of characters would reduce the accuracy of the method.

The method in [5] is based on LRA. In this method, the boundary for each character in the text line is fixed using boundary growing approach. The direction of the text line with respect to horizontal axis is obtained by growing the boundary till it reaches the pixel of neighboring character. The method considers all black pixels present in the document without segmenting individual text lines. The linear regression analysis is used to find slope of a skewed document using all pixel coordinate values. This method gives better accuracy up to 10° but fails when non-textual region is encountered in the document.

In [6], a method based on morphology is proposed, in which the image is dilated using a line structuring element of length 64 and later eroded by the same structuring element of length 512. Brent's method of parabolic interpolation is used for estimating skew. This method has mean absolute error of 0.2 and mean square error of 0.25.

Reference paper [7] has proposed skew detection method based on the Hough Transform (HT). The HT is computed at all angles of θ between 0 and 180 degrees. A heuristic measures the rate of change in accumulator values at each value of θ . The skew angle is set to the value of that maximizes the heuristic. These methods are computationally very expensive. In [8] only bottom pixels of candidate objects within a selected region are fed to Hough transform. The hierarchical Hough transformation technique is also adopted in another paper [9]. The main idea of the above methods is to reduce the amount of input data, but their computational complexities are still very high. In [10] they proposed an improved method to overcome the drawback of the method in [8]. Method in [12] considers some selected characters of the text selected characters which are subjected to thinning and Hough transform to estimate skew angle accurately.

As an attempt to reduce the computing cost and to gain high accuracy of the above Hough transform based approaches, two skew angle detection methods are proposed in this paper. The details of the proposed methods are provided in section II. In section III, the experimental results are presented and in section IV a comparative study of the some of the above mentioned HT based methods with the two proposed methods is presented. Conclusion is given in section V.

II. PROPOSED METHODS

This section presents the two proposed methods namely word centroid approach and dilate & thin approach. The first method is based on finding centroids of all words to estimate skew angle; it is explained in detail in Section A. The second method identifies each word as a single blob and finds the orientation of different blobs; this is explained in Section B.

A. Word Centroid Approach

This method is based on identification of centroid of selected words in the document image using connected component analysis and finding optimum skew angles of word centroids

using Hough transform. Initially, the text document is preprocessed by extracting prominent edges. Prewitt mask is used for edge detection because the input images are assumed to be free of noise and Prewitt mask is easy to implement. The extracted edges are then dilated for identification of words in the document image, as shown in Fig. 2. The dilation is an operation that grows or thickens object in binary image. The basic form of gray scale image dilation computes, for each image pixel, the maximum value of its neighboring pixels. The neighborhood is defined by the *structuring element*. Determining the size and shape of a structuring element is largely an empirical process. However, the overall selection of a structuring element depends upon the geometric shapes you are attempting to extract from the image data. For example, if you are dealing with biological or medical images, which contain few straight lines or sharp angles, a circular structuring element is an appropriate choice. When extracting shapes from geographic aerial images of a city, a square or rectangular element will allow you to extract angular features from the image. The size of the structuring element depends upon what features you wish to extract from the image. Larger structuring elements preserve larger features while smaller elements preserve the finer details of image features. We have used a square structuring element.

Later, connected component analysis is performed. For each component, coordinates of extreme (both in x and y directions) pixels are calculated and a bounding box is defined which determines the average height of all the components. The components whose height is very small are ignored. The word centroids of the selected characters as shown in Fig. 3 are extracted and fed to Hough transform.

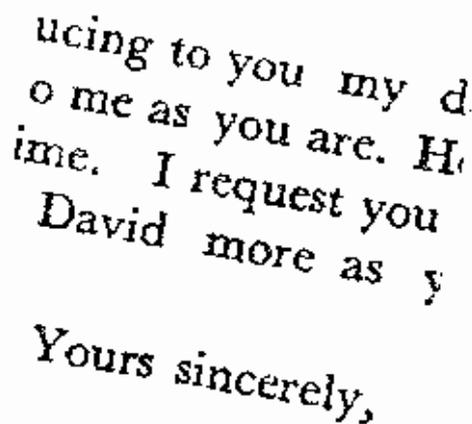


Fig. 1 Original image skewed at 10°

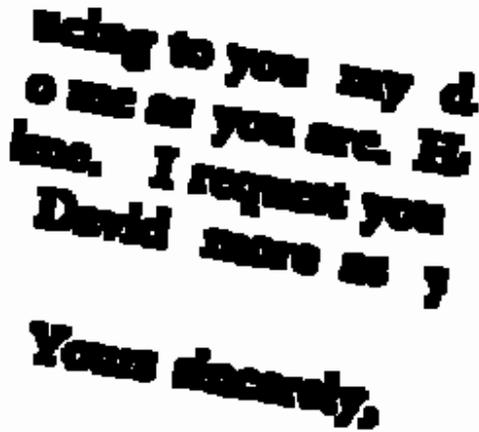


Fig. 2 Dilated image of Fig. 1 identifying the words

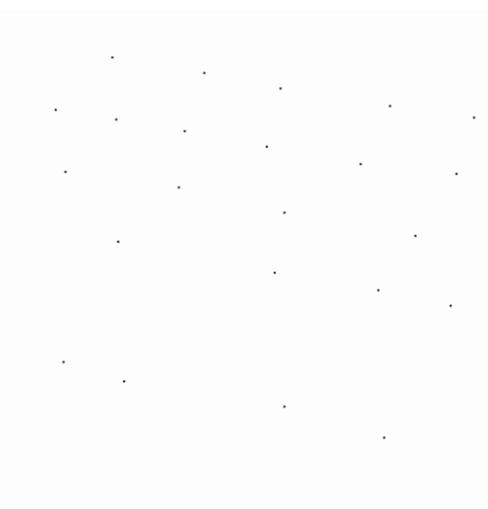


Fig. 3 Final coordinates fed to HT

B. Dilate & Thin Approach

This method has two stages. In the first stage, selected characters from the document image are blocked and dilated in order to obtain word blocks and then thinning is performed over the word blocked regions. A horizontal line structuring element is used for dilation. In the second stage, thinned coordinates are fed to Hough transform to estimate the skew angle accurately.

All of the connected components in a document image are identified by connected component analysis algorithm. For each component, coordinates of extreme (both in x and y directions) pixels are calculated and a bounding box is defined. The average height of the bounding box is determined. For each component whose height is more than average height, the bounding box height is reduced to average height. Also the components whose box height is very small are not considered, so that dots of the character like i and j, punctuation marks like full stop, comma, hyphen etc., are removed. The selected components in the document of fig.1 are blocked as shown in the Fig. 4.

Dilate the blocks with horizontal line structuring element until the character blocks of a word are merged as shown in Fig 5. Apply thinning algorithm to each block present in the document image. Thinning algorithm in [13] is used to thin the component blocks. The unique feature that distinguishes the thinning system is that it thins symbols to their central lines. This means that the shape of the symbols is preserved. It also means that the method is rotation invariant. The system has 20 rules in its inference engine. These rules are applied simultaneously to each pixel in the image. Therefore, the system has the advantages of symmetrical thinning and speed. The system is very efficient in preserving the topology of symbols and letters written in any language. Fig. 6 shows the result of thinning over the blocked component. Thus obtained points from Fig. 6 are then subjected to Hough transform to estimate the skew angle accurately. Detailed Hough transform is explained in section C.



Fig. 4 Selected components from Fig. 1



Fig. 5 Dilated components of Fig. 4

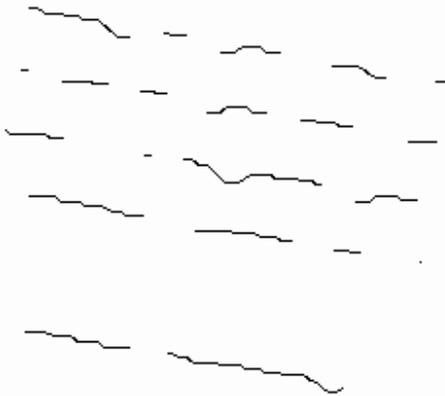


Fig. 6 Thinned word blocks from Fig. 5

C. Hough Transform

The Hough transform is widely used in image analysis, computer vision and digital image processing. It is a technique used to find shapes in a binary digital image. This approach is preferred when the objective is to find lines or curves in an image. It is defined by the parametric representation used to describe lines in the picture plane. It was introduced by Paul Hough in 1962 and patented by IBM. The transform parameterizes a description of a feature at any given location in the original image space. The basic idea is 'each straight line in an image can be described by an equation and each white point if considered in isolation could lie on an infinite number of straight lines. In the Hough transform each point votes for every line it could be on. The lines with the most votes win.'

In the image space, the straight line can be described as $y = mx + b$ and is plotted for each pair of values (x, y) . However, the characteristics of that straight line is not x or y , but its slope m and intercept b . Based on this fact, the straight line $y = mx + b$ can be represented as a point (b, m) in the parameter space (b vs. m graph.). Using slope-intercept parameters could make application complicated since both parameters are unbounded. As lines get more and more vertical, the magnitudes of m and b grow towards infinity. For computational purposes, however, it is better to parameterize the lines in the Hough transform with two other parameters, commonly called ρ (rho) and θ (theta). The parameter ρ represents the distance between the line and the origin, while θ is the angle of the vector from the origin to this closest point. Using this parameterization, the equation of the line can be written as: $\rho = x \cos \theta + y \sin \theta$. It is therefore possible to associate to each line of the image, a couple (ρ, θ) which is unique if $\theta \in [0, 2\pi]$ and $\rho > 0$. The (ρ, θ) plane is sometimes referred to as Hough space for the set of straight lines in two dimensions.

The Hough transform algorithm uses an array, called accumulator, to detect the existence of a line $y = mx + b$. The dimension of the accumulator is equal to the number of unknown parameters of the Hough transform problem. For example, the linear Hough transform problem has two

unknown parameters: m and b . The two dimensions of the accumulator array would correspond to quantized values for m and b . For each pixel and its neighborhood, the Hough transform algorithm determines if there is enough evidence of an edge at that pixel. If so, it will calculate the parameters of that line, and then look for the accumulator's cell that the parameters fall into, and increase the value of that cell. By finding the cells with the highest values, typically by looking for local maxima in the accumulator space, the most likely lines can be extracted, and their (approximate) geometric definitions read off. This is illustrated with an example in Figs. 7 (a) and 7 (b). For each data point, a number of lines are plotted going through it, all at different angles. These are shown here as solid lines. For each solid line, a line is plotted which is perpendicular to it and which intersects the origin. These are shown as dashed lines. The length and angle of each dashed line is measured. In the Fig. 7 (a), the results are shown in tables. This is repeated for each data point. A graph of length against angle, known as a Hough space graph, is then created Fig. 7 (b). The point where the lines intersect gives a distance and angle. This distance and angle indicate the line which bisects the points being tested.

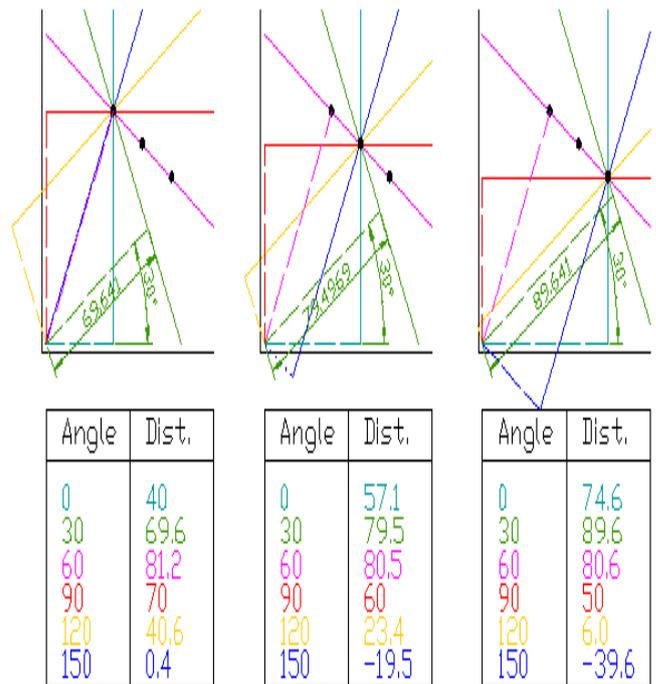


Fig. 7 (a) Lines plotted through points in an image along with their distance(ρ) and angle(θ)

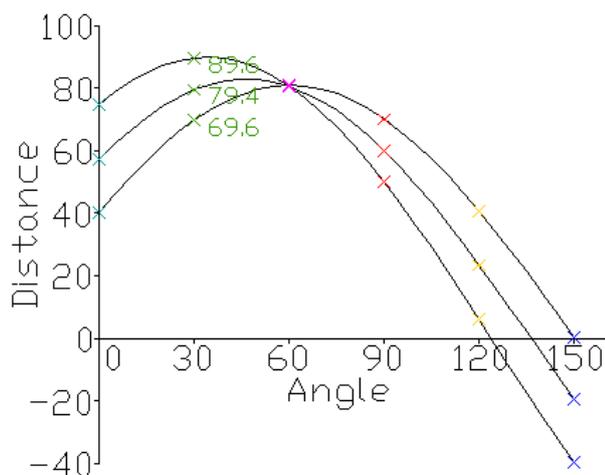


Fig. 7 (b) Hough space for lines in Fig. 7 (a)

III. EXPERIMENTAL RESULTS

This section presents the results of the experiments conducted to study the performance of the proposed method. The method has been implemented in the ANSI-C language on a Pentium IV 3GHZ. For experimentation purpose 20 documents are considered. The documents are tested with predefined angles that are varied between 3° and 20°. Mean Skew Angle (M) and Standard Deviation (SD) obtained by the proposed methodology for the 20 documents are reported in Table I.

TABLE I
 MEAN (M) AND STANDARD DEVIATION(SD)

True angle in degrees	Word Centroid approach		Dilate & Thin approach	
	M	SD	M	SD
3	3.075	0.3122	2.89	0.340
5	5.075	0.3240	4.90	0.357
10	9.825	0.3605	10.04	0.328
15	14.94	0.3427	14.91	0.293
20	19.87	0.3500	20.00	0.349

Fig. 8 shows the line graph for SD of the two proposed methods.

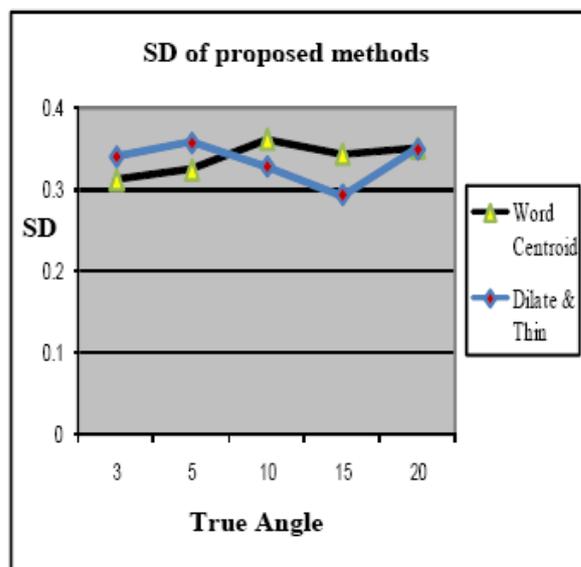


Fig. 8 The graph of SD of two proposed methods

IV. COMPARATIVE STUDY

A comparative study with three of the existing methods is carried out to establish the superiority of our method in terms of accuracy and efficiency. The 20 scanned text document image with different skew angles say 3, 5, 10, 15 and 20 degrees are actually considered as an input to the proposed method as well as to the existing methods. The mean and standard deviation obtained using the proposed method and the other methods are reported in Table II.

V. CONCLUSION

The experimental results for the proposed method are encouraging. The approach is robust for machine printed document containing only text. Further the method showed the same accuracy compared to other methods based on Hough transform. However, both the proposed methods fail for images containing pictures.

ACKNOWLEDGMENT

We thank Dr. Manjunatha Aradhya V N, Professor, Department of Information Science & Engineering, Dayanand Sagar College of Engineering, Bangalore for providing the test images.

TABLE II
MEAN (M) AND STANDARD DEVIATION (SD) OBTAINED FOR THE TESTED SKEW DOCUMENT IMAGES WITH DIFFERENT METHODS IN [7],[8] AND [10] COMPARED WITH THE SD OF OUR PROPOSED METHODS

True Angle in degrees	Srihari and Govindraju (1989)[7]		Le et al. (1994)[8]		Pal and Chaudhari (1996)[10]		Word Centroid approach		Dilate & Thin approach	
	M	SD	M	SD	M	SD	M	SD	M	SD
3	2.9	0.352	2.9	0.316	2.94	0.309	3.075	0.312	2.89	0.34
5	4.98	0.289	4.94	0.236	4.96	0.228	5.075	0.324	4.9	0.357
10	9.97	0.296	10.05	0.306	10.03	0.337	9.825	0.360	10.04	0.328
15	14.93	0.3	14.97	0.331	14.94	0.346	14.94	0.342	14.91	0.293
20	20.04	0.268	20.05	0.360	20.00	0.340	19.87	0.35	20.00	0.349

REFERENCES

- [1] Baird H. S., "The skew angle of printed documents", Proc. of SPSE 40th Symposium on Hybrid imaging systems, Rochester, NY, 1987, pp 739-743.
- [2] Postl W., "Detection of linear oblique structure and skew scan in digitized documents", Proc. of Int. Conf. on Pattern Recognition, 1986, pp 687-689.
- [3] Yan, H., "Skew correction of document images using interline cross-correlation", Computer Vision, Graphics, and Image Processing 55, 1993, pp 538-543.
- [4] A. Hashizume, P. S. Yeh, A. Rosenfeld, "A method of detecting the orientation of aligned components", Pattern Recognition Letters, 1996, pp. 125-132.
- [5] Shivakumara P., S. Guru, G. Hemantha Kumar, P Nagabhushan, "Skew detection in Binary document image using Linear Regression Analysis", proc. Of National Conf. on Advanced Computer Application NCAC-2002, Pollachi, India 2002, pp 41-46.
- [6] Najman L., "Using mathematical morphology for document skew estimation", SPIE Document Recognition and retrievals XI vol. 5296, 2004, pp 182-191.
- [7] Srihari S. N. and Govindraju V., "Analysis of textual images using Hough Transform", Machine vision Applications 2, 1989, pp 141-153.
- [8] Le D S, Thoma G R and Wechsler H, "Automatic page orientation and skew angle detection for binary document images." Patter Recognition 27, 1994, pp. 1325 – 1344.
- [9] B. Yu and A. K. Jain, "A robust and fast skew detection algorithm for generic documents," Pattern Recognition, 29, no. 10, 1996, pp. 1599-1630.
- [10] Pal U and Chaudhari B. B, "An improved document skew angle estimation technique", Pattern Recognition Letters, Vol. 17,1996, pp. 899-904.
- [11] B. V. Dhandra, V. S. Malemath, Mallikarjun H, Ravindra Hegadi, "Skew Detection in Binary Image Documents Based on Image Dilation and Region labeling Approach", The 18th International Conference on Pattern Recognition (ICPR'06), 2006.
- [12] Manjunath Aradhya V N, Hemantha Kumar G. and Shivakumara P, "Skew detection technique for binary document images based on Hough transform", International Journal of Information Technology, Vol. 3, 2006.
- [13] M Ahmed and R Ward, "Rotation Invariant Rule-Based Thinning Algorithm for Character Recognition", IEEE. Trans. Pattern Analysis and Machine Interlligence, vol. 24, No. 12, December 2002.
- [14] Gonzalez R., Woods, Digital Image Processing, Addison-Wesley Publishing Company. 2nd Ed. 2002.
- [15] D.R.Ramesh Babu Piyush M Kumat Mahesh D Dhannawat,"Skew Angle Estimation and Correction of Hand Written, Textual and Large areas of Non-Textual Document Images: A Novel Approach", IPCV 2006, 510-515.
- [16] Wikipedia - Hough transform source URL:http://en.wikipedia.org/wiki/Hough_transform