

Skew and Slant Correction for Document Images Using Gradient Direction

Changming Sun*

Deyi Si

CSIRO Math. & Information Sciences
Locked Bag 17, North Ryde
NSW 2113, AUSTRALIA

Dept. of Electronics Engineering
Beijing Univ. of Aeronautics & Astronautics
Beijing 100083, CHINA

Abstract

A fast algorithm is presented in this paper for skew and slant correction in printed document images. The algorithm employs only the gradient information. The skew angle is obtained by searching for a peak in the histogram of the gradient orientation of the input grey-level image. The skewness of the document is corrected by a rotation at such an angle. The slant of characters can also be detected using the same technique, and can be corrected by a shear operation. A second method for character slant correction by fitting parallelograms to the connected components is also described. Document images with different contents (tables, figures, and photos) have been tested for skew correction and the algorithm gives accurate results on all the test images, and the algorithm is very easy to implement.

Keywords: *Skew correction; Slant correction; Gradient orientation; Document image analysis.*

1 Introduction

There are many application areas in document image processing, including the selection of encoding methods for document archiving, retrieval, high-quality facsimile, and digital reprographics as well as preprocessors for OCR. The first step in document analysis is to acquire a digitized raster image of the document using a suitable scanning system. Then it is followed by page lay-out analysis and character recognition. Before the structure of the text is obtained, a test is carried out to find out whether the document is skewed. If it is skewed, correction needs to be done. Uncompensated skew can cause serious performance deterioration. Fig. 2(a) shows one of the skewed document images.

Many methods have been developed for the correction of skewed document images. Basically they can be described in five categories: 1) using projection profile; 2) using Hough transform technique; 3) Fourier method; 4) by nearest-neighbor clustering; and 5) correlation. In the projection profile method, a series of projection profiles are obtained at a number of angles close to the expected orientation, and the variation is calculated for each of the profiles. The profile that gives the maximum variation corresponds to the projection with the best alignment to the text lines. This projection angle is the skew angle. Baird proposed modifications to the projection profile method for very fast and accurate iterative convergence on the skew angle [1]. Spitz adapted Baird's alignment technique to fiducial point position data for detecting skew in compressed images [8]. Ciardiello *et al* have suggested a method using the projection histogram in which a sample region with the maximum average density of black pixels per row is rotated by pre-specified angles [2]. The horizontal projection histogram of the region is evaluated for each angle. The skew angle corresponds to a rotation for which the mean square deviation of the histogram is maximized. Nakano *et al* [6], Srihari and Govindaraju [9] and Hinds *et al* [5] proposed skew detection techniques based on the Hough transform. The Hough transform maps each point in the original (x, y) plane to all points in the (ρ, θ) Hough space of lines through (x, y) with slope θ and distance ρ from origin. The dominant lines are found from peaks in the Hough space and thus the skew. One limitation of this method is that if text is sparse, it may be difficult to choose a peak in the Hough space. Postl [7] proposed two methods for skew correction. One of them is called the "simulated skew scan" method, and it can be classified in the projection profile categories. The other method is based on the Fourier transform. In this method, the

*Email: changming.sun@cmis.csiro.au

direction for which the density of the Fourier space is the largest gives the skew angle. For large images, the Fourier method can be computationally expensive. And very often for a document image, the largest density direction of the Fourier space is on a vertical line and the true density direction may not be the largest! This makes the detection/search for the skewness of a document image difficult for the Fourier method. A bottom-up technique for skew estimation based on nearest-neighbor clustering is described by Hashizume *et al* [4]. In this work, the 1-nearest-neighbors of all connected components are found, the direction vectors for all nearest-neighbor pairs are accumulated in a histogram, and the histogram peak is found to obtain the document skew angle. This method has the advantage over the previous methods in that it is not limited to any range of skew angle. But since only one nearest-neighbor connection is made for each component, connections with noise, subparts of characters, and between-line connections can reduce the accuracy of the method. Yan [10] introduced a method for determining the skew angle of an image using cross-correlation between lines at a fixed distance. It is based on the observation that the correlation between two vertical lines in an image of a skewed document is maximized in general if one line is shifted relatively to the other line such that the character base line levels for the two lines are coincident.

In this paper we propose a simple and fast algorithm for determining the skew angle of an image and the slant angle of text characters only using the gradient orientation histogram. Its development is based on the observation that the gradient orientation should be mainly in the direction perpendicular to the text line. After the skew angle has been determined, the image can then be rotated for correction. Slant characters can also be detected using the same gradient orientation technique. A second slant detection algorithm will be proposed. The slant correction is performed by shearing the image in the direction of text line.

2 Skew Correction

For an image described by $f(x, y)$, the gradient vector $[p, q]^T$ at point (x, y) is defined by:

$$p = \partial f(x, y) / \partial x, \quad q = \partial f(x, y) / \partial y. \quad (1)$$

The orientation of this gradient vector is:

$$\phi = \arctan(q/p). \quad (2)$$

The gradient is obtained by using an $N \times N$ (in our experiment, $N=5$) Sobel filtering operation [3]. It splits the kernels into 2×1 sub-kernels and iteratively calculates the responses. The result is two responses for the

x - and y - directions respectively. Then the gradient orientation at this point of the image can be obtained using Eq. (2). The domain of ϕ is $[-\pi, \pi)$. For detecting the orientation of a skewed document, half of the range of ϕ will be enough. The negative values of ϕ are offset by π to become positive. If we have some knowledge about the range of the skew angle, we can reduce the domain of ϕ even further.

Our algorithm is based on an observation about gradient orientation distribution of the image. For a skewed document, there will be more points in the image whose gradient orientations are perpendicular to the document text lines. It is expected that the statistical information of the gradient orientation of an image can be used for skew angle detection. Fig. 1 gives an example shape of the gradient orientation histogram, in which θ is the skew angle. The orientation histogram of this gradient image can be obtained as described in Eq. (2). The angle ϕ in the equation is a continuous function. Quantization is needed for the given range and resolution. The resolution of the obtained histogram depends upon the range considered and the number of points in the histogram. For instance, if we would like to have 360 points within the angle between 45° and 135° , then the angle resolution will be 0.25 degree. From the obtained histogram $h(\phi)$, the orientation of the skewed document is the difference between angle ϕ where it gives the maximum value for $h(\phi)$ and $\pi/2$. For an upright document image, ϕ will be $\pi/2$, hence the skew angle is zero. In Fig. 2, (a) shows the original image of a skewed document, (b) gives the histogram of the gradient orientation of the original image.

After the skew angle of the document has been obtained, the input image can be rotated for skew correction. During the rotation operation, bilinear interpolation of neighbouring pixels was involved to reduce noise effect. The rotation can be realised using Eq. (3), where (c_x, c_y) is the centre of rotation, and r_{ij} , $1 \leq i, j \leq 2$, are the elements of the rotation matrix.

$$\begin{bmatrix} X_{out} \\ Y_{out} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} X_{in} \\ Y_{in} \end{bmatrix} - \begin{bmatrix} c_x \\ c_y \end{bmatrix} \quad (3)$$

Our proposed process for achieving the skew angle correction is:

1. Perform gradient operation on the image;
2. Obtain orientation histogram and smoothing it with a median filter;
3. Search for maximum in this histogram to obtain an initial skew angle;
4. Refine the obtained initial value by locally fitting a cubic polynomial function, and calculate

- the maximum analytically; and
 5. Rotate image for skew correction.

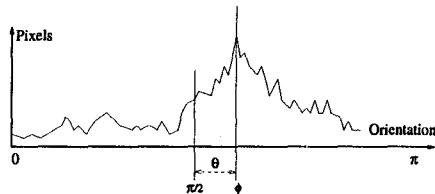


Fig. 1: An illustration for skew detection of a document image using gradient orientation histogram. The skew angle θ is the difference between $\pi/2$ and ϕ .

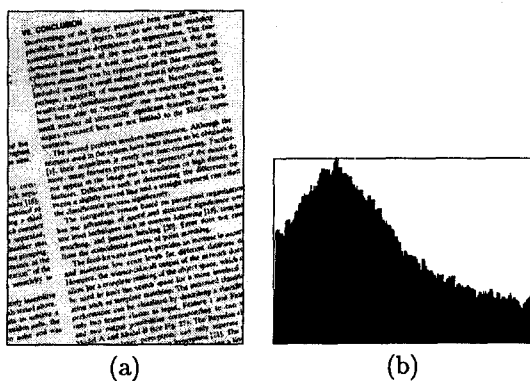


Fig. 2: (a) Original skewed document image; (b) Gradient orientation histogram of the original image in the range $[90-\phi_1, 90+\phi_2]$.

3 Slant Detection

Slant is the angle in degrees clockwise from vertical at which the characters were drawn. Slant correction is the process which tries to normalise the slant of the characters in a line or paragraph to the vertical. The same technique described in Section 2 can also be used for character slant correction by applying the technique to a line or a paragraph. After the slant angle has been detected, a shear operation on this text line or paragraph can be applied to correct the slant characters.

During the shear operation, for each pixel (i, j) in the original image, we compute the new coordinates (x, y) of this pixel in the slant-corrected image as follows:

$$\begin{cases} y = j \\ x = i - (\text{height} - j) * \tan(\theta) \end{cases} \quad (4)$$

where θ is the slant angle obtained. Point "A" is transformed into point "B" as in Fig. 3. Points on

the bottom row of the image will not be modified by the transformation. The height of the words (characters) were not changed while the width of the image will probably change. As a result of this transformation, the greater the value of the slant, the greater the variation of the aspect ratio of the image.

Fig. 4(a) shows lines of character with slant, and (b) gives the slant corrected image. Fig. 5 shows the slant correction for a grey level image. Fig. 5(a) is the original image showing the slanted word; (b) is the corresponding gradient orientation histogram; and (c) is the slant corrected image using the gradient method.

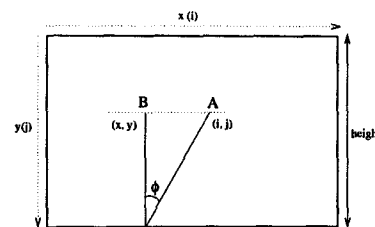


Fig. 3: Shear transformation of an image.

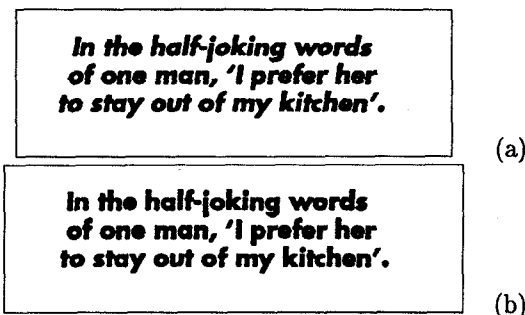


Fig. 4: (a) A few text lines with slant characters; (b) Slant corrected text (the slant angle is 8.33 degrees).

The character slant angle may also be detected by using the slant angle of the minimum bounding box of every connected component within the text line. This bounding box should not be a rectangle, but rather a parallelogram with top and bottom sides parallel to the text line. The mode of directions of these parallelograms relate to the slant angle. The slant angle is chosen as the median value of all the parallelogram angles considered. Fig. 6 shows the slant correction by fitting a parallelogram to each of the connected components of the binarized image. Fig. 6(a) is the original image showing the slanted word (binarized version of Fig. 6(a)); (b) shows the fitted parallelograms for each connected component; and (c) is the slant corrected image.

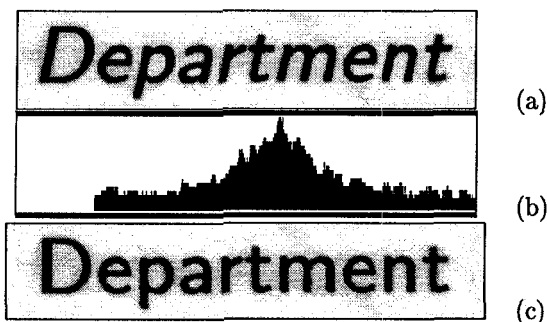


Fig. 5: (a) A word with slant characters (grey level image); (b) The gradient orientation histogram; (c) Slant corrected text (slant angle obtained is 12.04 degrees).

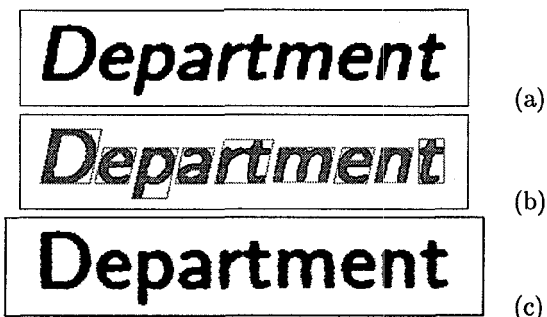


Fig. 6: (a) A word with slant characters; (b) Parallelograms fitted to the characters (connected components); (c) Slant corrected text (slant angle is 12.0 degrees).

4 Experimental Results

The results for the above described algorithm on some real images are given in this section. The gradient is obtained by using an 5×5 Sobel filtering operation on all the image points [3]. Then the gradient orientation at every point of the image can be obtained using Eq. (2). By having a set of bins in the range of $90 - \phi_1$ and $90 + \phi_2$, the orientation histogram of the input image can therefore be obtained. No initial smoothing was applied to the original image before the gradient operation, as the process of obtaining the histogram has the effect of canceling the random noise contribution. Fig. 7 shows the corrected images of Fig. 2(a). Most of the time of the algorithm was spent on the gradient operation, which can be done in hardware. The code has not been optimized for speed-up.

Because of the quantization effect, the text letter stroke is mostly zigzagged short line segments, and very often these line segments are either horizontal or vertical. Therefore, in most of the gradient orientation histograms, peaks often appear at 45° and its multi-

ples (such as 90° , 135° , 180° , 225° , 270° , and 315°). These distinct peaks have been removed by applying a median filter over the histogram.

The algorithm has been tested on various document images. They may only contain text, or may contain text with tables, graphics, or photographs. Fig. 8(a) shows one of the document images containing tables; (b) gives the skew corrected image. Fig. 9(a) illustrates an image with graphics; and (b) shows the skew corrected result. Fig. 10(a) gives one of the document image with photographs; and (b) is the skew corrected image.

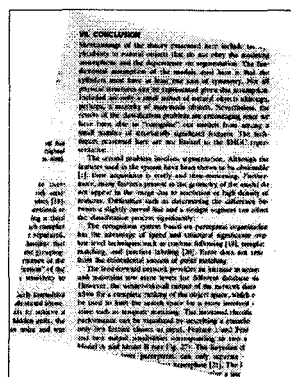


Fig. 7: Skew corrected image of Fig. 2(a).

5 Conclusions

A simple and fast document image skew correction algorithm has been developed which employs only the gradient information. The results show that the statistics of the gradient orientation are enough to obtain the skew angle of a document image. The same algorithm can also be used for the detection of the slant angle of the letters in a text line. The slant angle can also be detected by analyzing the connected components of each line or paragraph. Each connected component is fitted with a parallelogram. The peak of the parallelogram orientation histogram will be the slant angle. These slanted letters can be corrected using a shear operation to an upright position for the ease of recognition. From the nature of the algorithm, only "low-level" processing is involved, and therefore, it is possible to implement the algorithm in hardware for fast processing. The gradient operation can also be performed when the document is being scanned (or a single pass over the image). Therefore the overall process can be configured as pipeline processing.

If the document has a very large portion (say, over half of the image) of "italic" letters or cursive font, this method may give a systematic error. In this case,

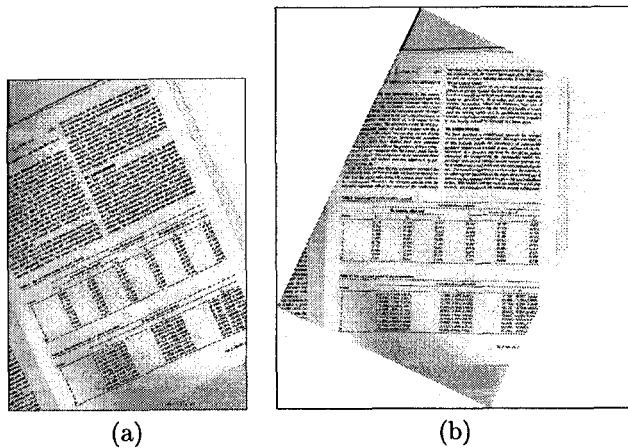


Fig. 8: (a) Original image with tables; (b) Skew corrected image (skew angle is 24.889 degrees).

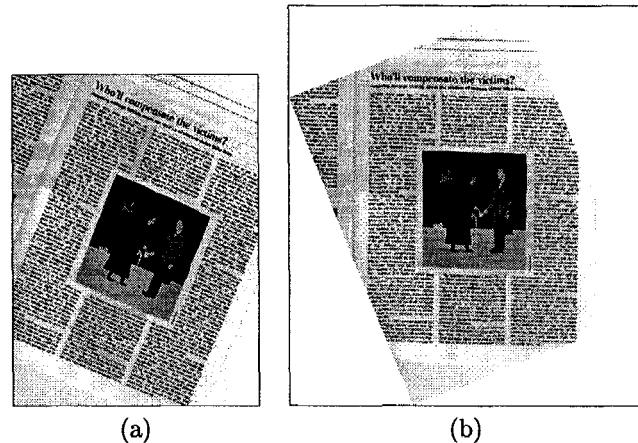


Fig. 10: (a) Original image with photographs; (b) Skew corrected image (skew angle is 22 degrees).

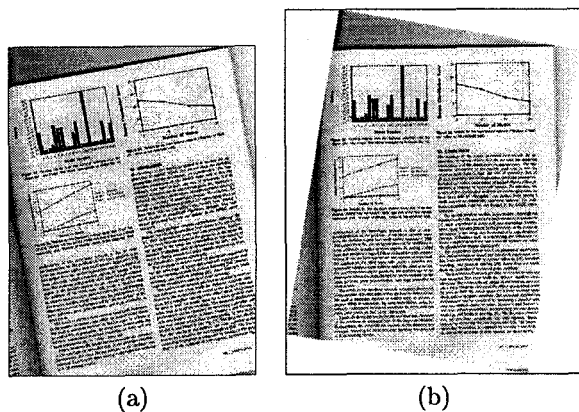


Fig. 9: (a) Original image with graphics; (b) Skew corrected image (skew angle is 9.33 degrees).

other techniques could be used. Future work is needed to compare our method with other techniques.

Acknowledgment

We thank the reviewers for their valuable comments.

References

- [1] H. S. Baird. The skew angle of printed documents. In *Proc. SPSE 40th Symp. Hybrid Imaging Systems, Rochester, NY*, pages 21–24, May 1987.
- [2] G. Ciardiello, M. T. Degrandi, M. P. Roccotelli, G. Scafuro, and M. R. Spada. An experimental system for office document handling and text recognition. In *Proceedings of International Conference on Pattern Recognition*, volume II, pages 739–743, 1988.

- [3] P. E. Danielsson and O. Seger. Generalized and separable Sobel operators. In H. Freeman, editor, *Machine Vision for Three-Dimensional Scenes*, pages 347–379. Academic Press, 1990.
- [4] A. Hashizume, P.-S. Yeh, and A. Rosenfeld. A method of detecting the orientation of aligned components. *Pattern Recognition Letters*, 4:125–132, 1986.
- [5] S. C. Hinds, J. L. Fisher, and D. P. D'Amato. A document skew detection method using run-length encoding and the Hough transform. In *Proceedings of International Conference on Pattern Recognition*, volume I, pages 464–468, 1990.
- [6] Y. Nakano, Y. Shima, H. Fujisawa, J. Higashino, and M. Fujinawa. An algorithm for the skew normalization of document image. In *Proceedings of International Conference on Pattern Recognition*, volume II, pages 8–13, 1990.
- [7] W. Postl. Detection of linear oblique structure and skew scan in digitized documents. In *Proceedings of International Conference on Pattern Recognition*, pages 687–689, 1986.
- [8] A. L. Spitz. Skew determination in CCITT Group 4 compressed document images. In *Proc. Symp. Document Analysis and Information Retrieval, Las Vegas, Nevada, USA*, pages 11–25, 1992.
- [9] S. N. Srihari and V. Govindaraju. Analysis of textual images using the Hough transform. *Machine Vision Applications*, 2:141–153, 1989.
- [10] H. Yan. Skew correction of document images using interline cross-correlation. *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, 55(6):538–543, November 1993.