An Accurate Method for Skew Determination in Document Images

S. Lowther, V. Chandran and S. Sridharan
Research Concentration in Speech, Audio and Video Technology
Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia
s.lowther@qut.edu.au, v.chandran@qut.edu.au, s.sridharan@qut.edu.au

Abstract

This paper presents a new Averaged Block Directional Spectrum (ABDS) technique for determining the skew angle of digitised document images. Skew determination is important because many other document analysis techniques require (near) perfectly aligned document images to work well. The technique is based on calculating the average 2D Fourier transform of blocks in a document image and using the Radon transform to find the peak in the directional spectrum. Results from experiments indicate that the new skew determination technique is highly accurate, obtaining the skew angle within 0.25° in 96.4% of test images. The technique has also shown robustness to documents highly corrupted by noise or having a large number of graphical elements.

1. Introduction

Skew Determination is an important tool for document analysis because many other document analysis techniques (such as OCR, script/language recognition and document segmentation) require (near) perfectly aligned document images to work well. If document skew is not corrected before further processing, the accuracy of these other techniques may be significantly reduced.

There exist a number of techniques that attempt to determine the skew angle of digitised document images [1], [2], [3], [4], [5] and [6]. Perhaps the most commonly used of these methods is that proposed by Baird [3]. There are commonly three problems associated with many skew determination techniques. The first such problem is that many techniques, such as those based on projection profiles as in [3], require a limited range of skew angles to be tested to avoid computational expense or reduction in accuracy. Another problem is that many techniques are designed to operate only on particular scripts and languages such as [5]. The last common problem with many skew determination techniques is the inability for many of them to accurately determine skew angle for documents containing various levels of graphics and/or images. With these problems in mind, a general method for skew determination across all possible angles, regardless of graphical content and script type is highly desirable as an initial pre-processing step in document analysis.

Techniques are proposed in [1] and [2] based on using the Fourier transform for accurate skew angle estimation. Postl [2] found that skew angle could be found from the 2D Fourier spectrum by either integrating radially or finding the angle of the highest valued peak. No mention was made on the presence of the dominant 'line' structure in Fourier space due to the line spacings of text, nor were results given. Peake and Tan [1] extended the earlier work of Postl by calculating skew angle from Fourier spectrums calculated from a number of equal sized blocks in the original document image. They noted that a directional alignment of energy occurs in Fourier space along an angle corresponding to the skew due to all line spacings and other harmonics of these line spacings. They also noted that values close the origin of the Fourier spectrum could be discarded to remove DC and very low spatial frequency components. Accuracies and typical parameters were not given.

2. Existing Algorithm

The skew determination technique by Peake and Tan used the dominant 'line' in Fourier space to determine skew according to the following algorithm [1]:

1. Image Subdivision – the image is divided into blocks of size NxN
2. Fourier Analysis – the Fourier spectrum is computed for each block (using FFT) and represented with the origin at the centre. A small window of size WxW, centred at the origin, is removed (set to zero) and the remaining values in the spectrum are normalised to obtain consistency of values between blocks.
3. Peak pair detection – the 5 highest pairs of peaks are located, and the corresponding angles calculated with respect to the vertical axis of the

spectrum. *The spectral peaks occur in pairs since the input is real-valued.*

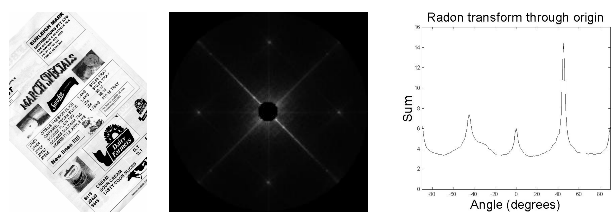
4. Histogram – a skew angle histogram is constructed where the bins are integer values of angles. For each angle calculated, the bin corresponding to its integer value is incremented by the normalised value of the peak in the Fourier spectrum. The real valued angles are retained.
5. Integer valued angle calculation – the histogram is then smoothed and its global peak selected as the integer value of the skew angle.
6. Real value angle calculation – all real valued angles within ±t° of the integer valued angle are selected and their median calculated. This median value represents the final skew angle of the document. The image may then be skew corrected through rotation (using bilinear interpolation) by the negative of the calculated angle.

Empirically we found N = 256, W = 25 and t = 5° gave good results on document images scanned at 150dpi.

3. ABDS Algorithm

Whilst Peake and Tan simply take the maximum values of the spectrum to calculate skew angle, experimental evidence has suggested that better accuracy can be found by applying radial integration on the average of Fourier spectrums taken from blocks in the document image. The full Fourier spectrum (as opposed to one half plane as in [2]) is used to achieve a higher angular resolution. A donut shaped mask is applied to the averaged Fourier spectrum to remove DC and very low spatial frequency components and make the radial bands of uniform bandwidth in all directions. By using a Radon transform modified to only find projections passing through the origin, the document orientation can be accurately determined by simply selecting the angle corresponding to the maximum value in the Radon transform. Speed and accuracy are combined by using a coarse Radon transform to first select the correct skew angle within a few degrees and then applying a fine Radon transform in smaller angle increments to determine skew within a defined accuracy. Since Peake and Tan found that increased accuracy is obtained by using binary document images, documents should first be thresholded using a simple global thresholding technique. The algorithm used for skew determination is presented below. Figure 1 shows example output of some of the steps of the proposed technique.

1. Image Subdivision – the image is divided into blocks of size NxN
2. Fourier Analysis – the full Fourier spectrum is computed for each block (using FFT) and represented with the origin at the centre. In this algorithm the window at the origin is not removed. All values in the Fourier spectrum are normalised to the range [0, 1] to obtain consistency between blocks.
3. Mean Fourier Calculation – The Fourier transform for all blocks is averaged together to form a single Fourier block for applying the Radon transform.
4. Apply a Mask – A donut shaped mask is applied to the mean Fourier block to remove DC and very low spatial frequency components and to make the radial bands of uniform bandwidth in all directions.
5. Approximate Angle Calculation – Apply a modified Radon transform to the mean Fourier block in the range [-90°, 90°] in angle increments of 1º. The peak of the Radon transformation is taken as the approximate angle of skew.
6. Refined Angle Calculation – Apply a second modified Radon transform to the mean Fourier block in the range ±t° of the approximate skew angle, in smaller angle increments.



(a) (b) (c)

Figure 1. (a) Example of a document with approximately 45° skew. (b) The mean Fourier transform of blocks taken from (a). (c) The first modified Radon transformation of (b) showing the major peak at approximately 45° and a subdominant peak at approximately -45°.

The value of N should be chosen as to obtain suitable angular resolution in the Fourier spectrum while minimising computational complexity of the Fourier transforms. If the chosen value for N is too small, the blocks may contain too few lines of text to accurately find the angle of skew while if N is too large, there are fewer blocks for averaging and fewer blocks containing only text, leading to a decrease in robustness. The smaller angle increments in step 6 in the above algorithm should be selected based on the desired accuracy and speed of the intended application. Empirically we found N = 256, t =

5° and smaller angle increments of 0.1° gave good results on document images scanned at 150dpi.

The process of averaging the Fourier blocks together has two main advantages. Firstly, it provides robustness from domination by any one value in the Fourier blocks, some of which may not contain lines of text or other horizontally aligned content, and secondly, it speeds up the skew determination algorithm since the Radon transforms only need to be applied to a single Fourier block.

Comparisons of the results found using the technique proposed in [1] compared to those of the ABDS technique are presented in the results section.

4. Resolution Accuracy Analysis

A limitation of the Fourier block in the proposed technique is its angular resolution which governs the accuracy of the detected skew angle. A higher angular resolution will allow the technique to detect a skew angle nearer to the true skew angle. Considering document images scanned at 150dpi, written in 12 pt font with single line spacing, the length between text lines is approximately 30 pixels. Given a Fourier block size of 256x256 with a resolution of 1/256 cycles/pixel and a peak line spacing frequency of 1/30 cycles/pixel, then the peak value in each Fourier block due to the line spacing will lie approximately 256/30 ≈ 9 pixels from the origin. The angular resolution of the peak value of the Fourier block is approximated by,

$$|d\theta| = \left|\frac{\partial\theta}{\partial x}dx + \frac{\partial\theta}{\partial y}dy\right|$$

using $\theta = \tan^{-1}\left(\frac{y}{x}\right)$, we have

$$|d\theta| = \left|\frac{-y}{x^2+y^2}dx + \frac{x}{1+x^2}dy\right|$$

or equivalently,

$$|d\theta| = \left|-\frac{\sin(\theta)}{r}dx + \frac{\cos(\theta)}{r}dy\right|$$

Therefore, at 9 pixels from the origin of the Fourier block along the vertical axis (r=9/256, θ=0˚, dx = 0, dy = 1/256), we have dθ ≈ 6.4º. Similarly, on the 45º diagonal to the vertical axis (r=30/256, θ=45˚, dx = 1/512, dy = 1/512), we get dθ ≈ 4.5º. These values give angular resolution in the range [4.5º, 6.4º] for finding the peak value of the Fourier blocks corresponding to the contribution of the line spacing in the original document. *By investigation of the Fourier block (as seen in Figure 1(b)), it can be seen that the dominant line extends much further than the peak value caused by particular line spacings. This dominant line is predominantly caused by harmonics of the inter line spacings and leakage and is dependant on the line spacings and the number of lines of text in the document.* The existence of the dominant line aids in achieving a finer angular resolution since points further from the origin than the peak Fourier value will contribute to the detection of the skew angle. If the dominant line contributes up to four times the length of the peak value (36 pixels from the origin), the angular resolution is calculated to be in the range [1.1º, 1.6º]. The results presented in the following section indicate that the practical angular resolution obtained by the use of our method is less than 0.25º. This higher angular resolution can be attributed to a higher contribution of the dominant directional energy along a line rather than only from the position of the dominant peak and also partly to the Radon transform computation. The Radon transform effectively sums up 'leaked' and 'harmonic' evidence and can be applied for the peak angle detection at better than Fourier resolution in this case.

5. Results

Tests have been conducted using both the ABDS technique and the technique in [1] across a range of documents and skew angles. The test set consisted of 94 document images randomly rotated 20 times each in the range [-45°, 45°] for a total of 1880 individual tests. The documents used consisted of various invoices, letters and billing statements written predominantly in English and containing varying levels of graphical content and line spacing. Each of the 95 documents was scanned at 150dpi and skew corrected manually before tests were conducted. Tests were conducted on the original document images and binary copies with a suitable binary threshold selected

Table 1. Results of testing the two skew determination techniques on binary document images against different angle accuracy thresholds. The percentages shown in the table correspond to the percentage of documents whose skew angle was correctly determined within the given error threshold.

Error Threshold	≤ 1º	≤ 0.5º	≤ 0.25º	≤ 0.125º
ABDS	96.6%	96.5%	96.4%	93.45%
Peake and Tan	74.15%	65.45%	48.5%	29.85%

using [7]. The results from testing on binary and greyscale documents are presented in Figures 2 and 3 respectively. For clarification, accuracies on binary images for key skew error thresholds are also presented in Table 1. Figures 4 and 5 present the distribution of absolute skew errors greater than 1°.

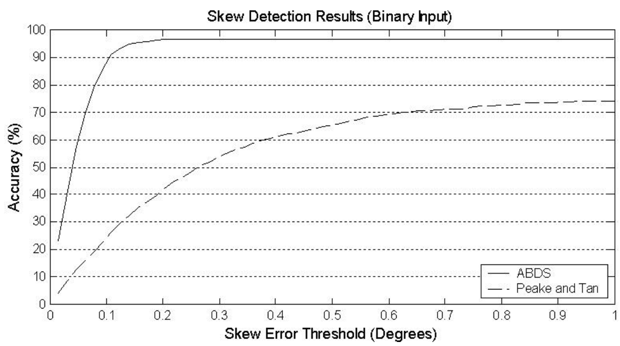


Figure 2. Results of testing the two skew determination techniques on binary document images. The graph shows the percentage of images for which error in skew determination was found within the given skew error thresholds.

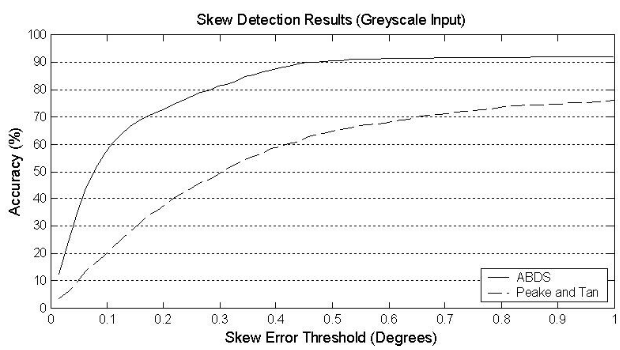


Figure 3. Results of testing the two skew determination techniques on greyscale document images. The graph shows the percentage of images for which error in skew determination was found within the given skew error thresholds.

Table 2. Accuracies obtained from testing on images with varying levels of graphical content using the ABDS technique for skew determination error within 0.25°.

Text Level (%)	**0-20**	**20-40**	**40-60**	**60-80**	**80-100**
Accuracy (%)	94.29	93.08	99.33	100	100

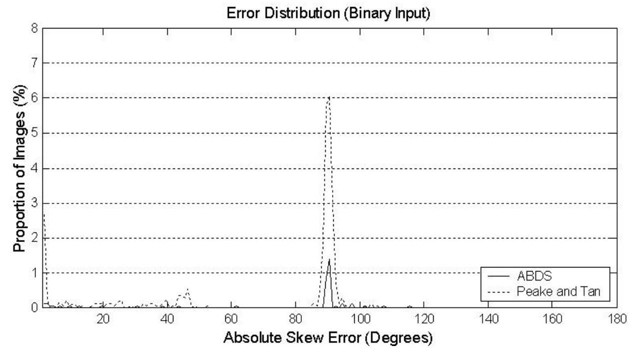


Figure 4. Skew error distribution for absolute skew error greater than 1° on binary images.

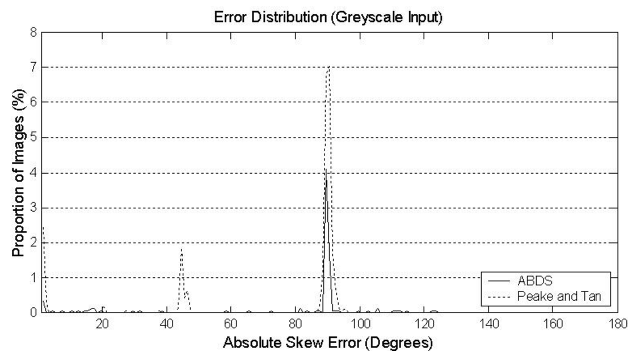


Figure 5. Skew error distribution for absolute skew error greater than 1° on greyscale images.

Tests were also conducted to evaluate the robustness of the ABDS technique to varying levels of graphical content. Test images were separated into classes according to the proportion of text pixels compared to non-text pixels in the binary copies of the original document images. Graphical content in the test images consisted mainly of logos, tables, graphics and borders oriented in the same direction as document text. Text pixels were found automatically using a technique similar to the one presented in [8]. Once the test images were separated, each binary image was randomly rotated 10 times in the range [-45°, 45°] and the corresponding skew determined using the ABDS technique. Accuracies were obtained from the proportion of test images for which skew determination error was within 0.25°. The results from testing can be seen in Table 2.

6. Discussion

As can be seen from Figures 2 and 3, the ABDS technique has performed significantly better than the technique in [1] for all angle accuracy thresholds. Figures

4 and 5 indicate that the majority of errors from use of both techniques occurred from the detection of the line at 90º to the true skew angle caused from character spacings, word spacings and graphical elements. Other techniques may be employed to determine the orientation of a document in these instances.

Table 2 shows that the ABDS technique is significantly robust to the presence of graphics and other non-textual elements in document images. While the ABDS technique produces more accurate results when there are a high proportion of textual regions in a document image, tests indicate that accuracies well above 90% for skew error within 0.25° can be obtained even if there is very little text present in the images. High accuracies in low text situations can be attributed to non-textual components in document images having similar orientation to that of textual components, therefore contributing to the dominant line in the Fourier spectrum.

The computational complexities for the calculation of the Fourier spectrum and the Radon transforms when the number of blocks and number of angles of projection are held constant are in the order of $N^2 \log_2 N$ and N respectively where N equals the width of the blocks used. Clearly the calculation of the Fourier spectrum dominates the computational complexity of the ABDS technique. As stated in [1], the Fourier spectra of the image blocks can be computed in parallel, increasing the throughput of the algorithm. Other speed refinements can include reduction of the block size and the range and number of angles investigated with the Radon transform calculations.

Further research will investigate the use of the new technique over a large range of script and document types as well as the suitability of the technique to low resolution images. A fast technique for removing the 90° ambiguity in the orientation of a deskewed document will also be investigated to ensure that a final algorithm can accurately estimate any possible skew angle regardless of the original document orientation.

7. Conclusion

In conclusion, we have built upon previous work in [1] to provide a new technique for determining the skew angle of digitised document images. The results from testing have shown that the technique is highly accurate, achieving a skew angle determination accuracy of 96.4% within 0.25º of the true skew angle for binary document images. While best results are obtained by use of binary document images, good results are still obtained on original greyscale images. The advantage of the technique over most other techniques is the ease of detecting skew over all possible angles as opposed to those within a particular range of angles and the robustness of the technique to various levels of graphical content.

7. Acknowledgements

The authors would like to thank the Defence Science and Technology Organisation (DSTO) for their support in our document analysis research.

8. References

[1] G. S. Peake and T. N. Tan, "A General Algorithm for Document Skew Angle Estimation", presented at Proceedings of the 1997 International Conference on Image Processing, pp. 230-233, 1997.

[2] W. Postl, "Detection of Linear Oblique Structures and Skew Scan in Digitized Documents", presented at International Conference on Pattern Recognition, pp. 687-689, 1986.

[3] H. Baird, S, "The Skew Angle of Printed Documents", in *Document Image Analysis*, L. O'Gorman and R. Kasturi, Eds.: IEEE Computer Society Press, pp. 204-208, 1995.

[4] R. Smith, "A Simple and Efficient Skew Detection Algorithm via Text Row Accumulation", presented at Proceedings of the International Conference on Document Analysis and Recognition, pp. 1145-1148, 1995.

[5] B. B. Chaudhuri and U. Pal, "Skew Angle Detection of Digitized Indian Script Documents", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 703-712, 1997.

[6] T. Steinherz, N. Intrator, and E. Rivlin, "Skew Detection via Principal Components Analysis", presented at Proc. ICDAR'99, Bangalore, India, pp. 153-156, 1999.

[7] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-9 (1), pp. 62-66, 1979.

[8] V. Wu, R. Manmatha, and E. M. Riseman, "Finding Text In Images", presented at Proceedings of the 2nd ACM International Conference on Digital Libraries, pp. 3-12, 1997.