# A Fast High Precision Algorithm for the Estimation of Skew Angle Using Moments

George Kapogiannopoulos
Department of Informatics and Telecommunications
University of Athens
Division of Communications and Signal Processing
TYPA Buildings, 157 71 Athens
Greece
email: gkapog@di.uoa.gr

Nicholas Kalouptsidis
Department of Informatics and Telecommunications
University of Athens
Division of Communications and Signal Processing
TYPA Buildings, 157 71 Athens
Greece
email: kalou@di.uoa.gr

## ABSTRACT

An efficient, high precision algorithm is proposed for the correction of the skew angle of scanned document images used in OCR (Optical Character Recognition). The algorithm employs moments to find the primary axis of every object in the document rather than the Hough transform and handles arbitrary angles. The object can be a halftone image or a line of text. Lines are described using the Freeman's chain code. A simple technique used to find the primary axis of each object from the chain code. Using a feature that depends on the size of the object, the weighted average angle is estimated. The accuracy is better than 0.02 degrees and the speed is high. The image is rotated in the opposite direction to get a skew corrected image.

**KEY WORDS:** OCR, skew angle, chain code, moments

## 1. INTRODUCTION

The need for automatic character recognition of printed documents is increasing because a huge amount of printed documents requires processing for updating databases, document archiving, automated document delivery and other applications. Recognition usually is done off-line, that means, the printed documents are scanned, stored as images and at later stage the character recognition takes place. During scanning operation some papers may not have been fed properly in the scanner causing skew angle to appear in the scanned images. In addition, some documents are not printed with acceptable quality so each column of text may have a small skew angle. If the original document is not available to rescan the image, -which causes unduly delays-, we get poor recognition results because most of the recognition processes expect documents with straight lines and not skewed.

Most existing approaches use the Hough transform or enhanced versions. The Hough transform detects straight lines in an image. The algorithm transforms each of the edge pixels in an image space into a curve in a parametric space. Suppose that N edge pixels $\{(x_0, y_0), (x_1, y_1), ..., (x_{N-1}, y_{N-1})\}$ are given and we want to find a set of straight lines that best fits them. An edge at location $(x_i, y_i)$ is transformed in the $(r, \theta)$ plane by:

$$r = x_i \cos\theta + y_i \sin\theta$$

If $n$ levels quantization is applied then the time complexity for the algorithm is O($nN$) which is too slow for real time applications. Good accuracy of the algorithm requires large values for $n$.

Many variants have been reported with the aim to reduce complexity [1] and increase accuracy. Hinds et al. [2] have proposed a scheme that uses the Hough transform with reduced input data. An intermediate image is created by placing the length of the vertical run length in its bottom-most pixel. The skew angle is estimated by applying the Hough transform on the intermediate image.

Another approach [3] based on the correlation between the vertical lines and the skew angle can be determined from the optimal shift and the distance between each pair of lines. The algorithm can be seen as a special case of the Hough transform. It is accurate for skew angles less than ±10 degrees.

A different technique based on projection profiles is described in [4]. The skew angle is obtained by calculating the arctangent of phase shift between projection profiles of adjoining equal sized document wide swaths. The algorithm is accurate within the range of ±10 degrees of skew. Another algorithm [5] reduces the parameter space in one dimension and works on histograms based on connected components rather than on pixels in steps of e.g. 0.5 degrees. The energy is calculated for each histogram and the maximum energy value specifies the skew angle.

The method presented in [6] estimates the skew angle from the centres of character boxes using a least square line fitting technique. Performance depends on the reliability of the character extraction procedure.

One technique which works in landscape oriented printed documents [7] uses the Hough transform in a reduced set of input data. This reduced set contains only the bottom pixels of the candidate objects. Again here the skew angle is in the range of ±15 degrees with accuracy of 0.5 degrees.

Another method [10] uses horizontal projections for a predefined set of angles in the [-10, +10] degrees range with a step of 0.2 degrees and a set of codebooks. These codebooks contain one vector for every rotation angle, where each vector is the KL-transform PSD (Power Spectrum Density) of the horizontal projection of that angle. The method is accurate enough for up to ±3.5 degrees and gives good results for an error tolerance of 1.2 degrees.

Almost all methods outlined above work in a limited angle range, have limited angle accuracy, and need a lot of input data and computation time. The algorithm proposed in this paper is simple, fast and works with any kind of objects, like bend lines (not only straight lines), pictures, columns of text etc. The algorithm is based on the idea that in order to find the skew angle of a line, the line is viewed as a solid object. The orientation $\theta$ of this object between the principal axis and the horizontal axis gives an exact estimation of the skew angle.

## 2. PRELIMINARIES

A binary image is a two-dimensional binary function whose values are 0 or 1, $i$ is the horizontal coordinate and $j$ is the vertical coordinate.

Smearing is used for closing white gaps in one direction vertical or horizontal. It works with the simple assumption that if a white gap is less than a certain limit then the white gap closes, which means that black pixels are inserted.

In chain coding the direction vectors between successive boundary pixels are encoded. Usually chain code employs four directions, when we consider 4-connected paths, which can be coded in 2-bit code words. It is clear that the chain code depends on the initial point of boundary following, but this is not a problem because the chain code can be considered as a circular list in the case of closed boundaries. In the case of 4-connected chain code [figure 1], segments having code 0 (or 2) contribute to object width, whereas segments having code 1 (or 3) contribute to object height [8]. For example the chain code 1,1,1,2,2,2,3,3,3,0,0,0 gives a rectangle with all the sides of length 3.

Pre-processing is necessary for the separation between pictures and text but also for the representation of the text in paragraphs and lines. We do not use pre-processing to separate the pictures from the text as it has no affect on the proposed skew angle estimation method, but only to separate the lines of the text.

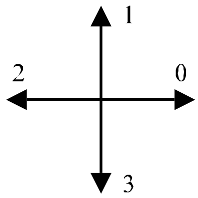
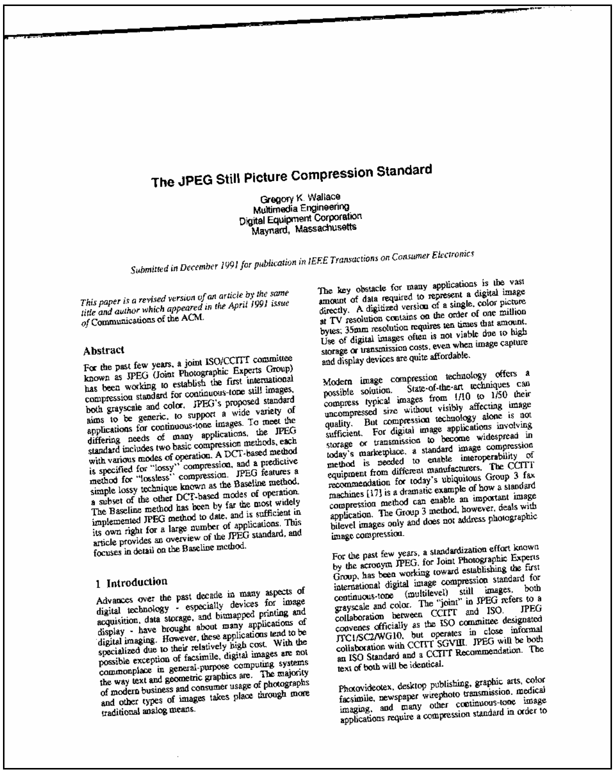



**Figure 1.** The direction vectors corresponding to a 4-connected chain code.



The JPEG Still Picture Compression Standard

Gregory K. Wallace
Multimedia Engineering
Digital Equipment Corporation
Maynard, Massachusetts

*Submitted in December 1991 for publication in IEEE Transactions on Consumer Electronics*

*This paper is a revised version of an article by the same title and author which appeared in the April 1991 issue of* Communications of the ACM.

**Abstract**

For the past few years, a joint ISO/CCITT committee known as JPEG (Joint Photographic Experts Group) has been working to establish the first international compression standard for continuous-tone still images, both grayscale and color. JPEG's proposed standard aims to be generic, to support a wide variety of applications for continuous-tone images. To meet the differing needs of many applications, the JPEG standard includes two basic compression methods, each with various modes of operation. A DCT-based method is specified for "lossy" compression, and a predictive method for "lossless" compression. JPEG features a simple lossy technique known as the Baseline method, a subset of the other DCT-based modes of operation. The Baseline method has been by far the most widely implemented JPEG method to date, and is sufficient in its own right for a large number of applications. This article provides an overview of the JPEG standard, and focuses in detail on the Baseline method.

**1 Introduction**

Advances over the past decade in many aspects of digital technology - especially devices for image acquisition, data storage, and bitmapped printing and display - have brought about many applications of digital imaging. However, these applications tend to be specialized due to their relatively high cost. With the possible exception of facsimile, digital images are not commonplace in general-purpose computing systems the way text and geometric graphics are. The majority of modern business and consumer usage of photographs and other types of images takes place through more traditional analog means.

The key obstacle for many applications is the vast amount of data required to represent a digital image directly. A digitized version of a single, color picture at TV resolution contains on the order of one million bytes; 35mm resolution requires ten times that amount. Use of digital images often is not viable due to high storage or transmission costs, even when image capture and display devices are quite affordable.

Modern image compression technology offers a possible solution. State-of-the-art techniques can compress typical images from 1/10 to 1/50 their uncompressed size without visibly affecting image quality. But compression technology alone is not sufficient. For digital image applications involving storage or transmission to become widespread in today's marketplace, a standard image compression method is needed to enable interoperability of equipment from different manufacturers. The CCITT recommendation for today's ubiquitous Group 3 fax machines [17] is a dramatic example of how a standard compression method can enable an important image application. The Group 3 method, however, deals with bilevel images only and does not address photographic image compression.

For the past few years, a standardization effort known by the acronym JPEG, for Joint Photographic Experts Group, has been working toward establishing the first international digital image compression standard for continuous-tone (multilevel) still images, both grayscale and color. The "joint" in JPEG refers to a collaboration between CCITT and ISO. JPEG convenes officially as the ISO committee designated JTC1/SC2/WG10, but operates in close informal collaboration with CCITT SGVIII. JPEG will be both an ISO Standard and a CCITT Recommendation. The text of both will be identical.

Photovideotex, desktop publishing, graphic arts, color facsimile, newspaper wirephoto transmission, medical imaging, and many other continuous-tone image applications require a compression standard in order to

**Figure 2.** Scanned document with a skew angle of about 3 degrees.

Our method uses Freeman's chain code to describe every object moving clockwise. The binary image is smeared horizontally so each line is becoming a solid object and the baseline is smoothed. It is obvious (if we no pictures appear in the document) that the number of objects will be about the number of lines in the document. The smear parameters depend only on the resolution of the scanned image and are automatically set, as TIFF files that are used in scanners contain the resolution information in dpi (dots per inch). Typical parameter for the horizontal smearing is the 1/60 of the image dpi. Thus in a 600 dpi image a gap of 10 or less white pixels is closed.

Freeman's Chain Code is used to encode every smeared object using 4 directions. In order to speed things up encoding uses the byte values of the image rather than the pixel values. After the smearing process is completed the outline of the object is smoothed so we can use the byte (8 pixels) information in the horizontal direction of the image. This means that a neighbour byte with all eight bits is encoded to the chain. Experiments show that this does not have any significant affect in the estimation of skew angle but reduces pre-processing by a factor of eight.

## 3. SKEW ANGLE ESTIMATION

The pre-processing of the previous section transforms the scanned page into a number of objects described by a chain code [figure 3]. Every such object is solid and is considered solid even if it has gaps consisting of very few pixels. Considering that a text line has a height of about the average height of the characters that constitute the line, the ratio of width/height is large. This implies that the direction of the line object gives us an estimation of the skew angle. A simple way to find the direction of an object, in our case a text line, is by using moments.

The moments of discrete images are given by:

$$m_{pq} = \sum_i \sum_j i^p j^q f(i,j)$$

$$\mu_{pq} = \sum_i \sum_j (i-\bar{x})^p (i-\bar{y})^q f(i,j)$$

If the image is binary the moments become:

$$m_{pq} = \sum_i \sum_j i^p j^q$$

$$\mu_{pq} = \sum_i \sum_j (i-\bar{x})^p (i-\bar{y})^q$$

Some important image features are obtained from the moments. The area of the object is given by the moment $m_{00}$. The centre of gravity is given by:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

Object orientation $\theta$ is the angle between the major axis of the object and the horizontal axis. It can be estimated by minimizing the function:

$$S(\theta) = \sum_i \sum_j [(i-\bar{x})\cos(\theta) - (j-\bar{y})\sin(\theta)]^2$$

The minimization gives the following value:

$$\theta = \frac{1}{2}\arctan\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right)$$

In our implementation we do not use the binary image of an object to compute the above features; instead we use the chain code to describe each object and compute the features. It can be shown that if the object is considered solid then the features such as the centre of mass and object orientation do not change if we calculate the moments using only the boundary pixels of the chain code representation. This leads to a reduction of computational complexity from $O(N^2)$ to $O(N)$.

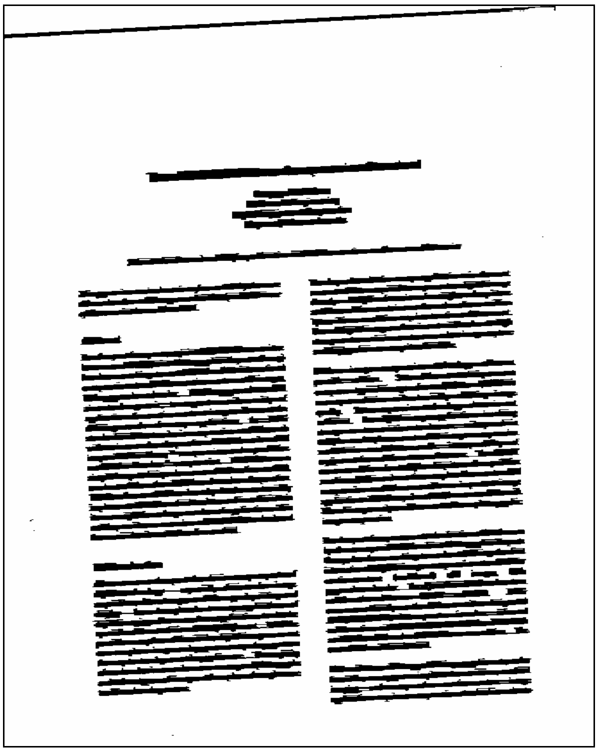


**Figure 3.** The scanned image after smear processing.

For each object described in chain code we retain the starting coordinates. Then for every code in the encoded chain we increase or decrease the coordinates accordingly. For faster implementation the chain code in the horizontal direction is described using the byte information of the image. For the calculation of the moments it is necessary to expand this byte to its own eight pixels in order to get more accurate object orientation.

We estimate the object orientation $\theta$ for each of the $m$ objects and the skew is estimated by the following formula:

$$skew = \frac{\sum_{i=1}^{m} \theta_i m_{00}^i}{\sum_{i=1}^{m} m_{00}^i}$$

Thus small objects consisting of few characters are not too wide and the object orientation $\theta$ does not exactly follow the baseline of the object. It is obvious that these objects have a small ratio of width/height and do not contribute to the skew estimation in the same manner as the objects consisting of a full line and have large width/height ratio. The number of the boundary pixels each object has is employed for weighting. It is obvious that moment $m_{00}^i$ provides the number of the boundary pixels of object *i*. This gives an approximation of the object size which is analogous to the ratio width/height, since the height is about the same for all the objects.

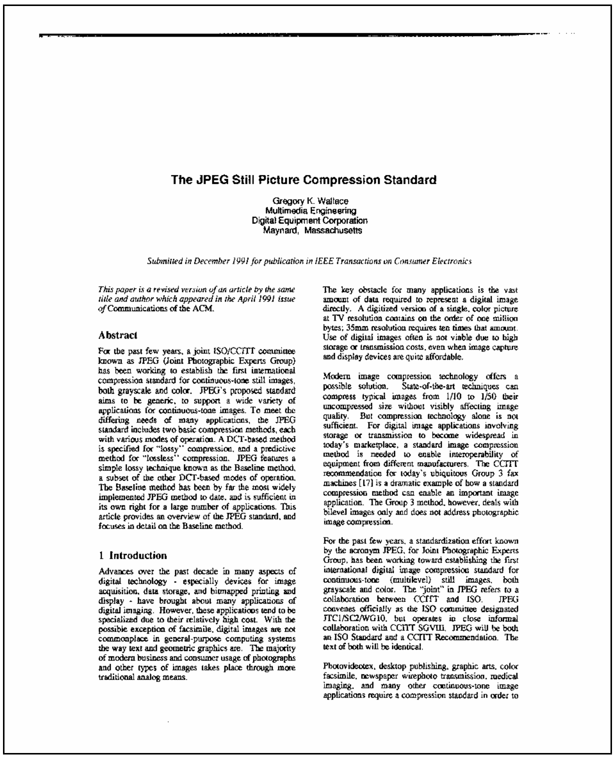
Objects like graphics whose estimated skew angles are exceeding a certain limit (1 degree) from the mean estimated skew angle, are not included in the estimation. This implies that images may have a radically different orientation that depends on the content and the width/height ratio of the image. However, the images can be removed from the page after proper pre-processing, but this is a problem which will be dealt with [11] in forthcoming research.

## 4. EXPERIMENTAL RESULTS AND CALIBRATION

Most images are typical A4 size of 8.5x11 inches and are scanned at a resolution of 300 dpi or 600 dpi. The computer used was a personal computer running at 1400 Mhz. The total time taken for the pre-processing and skew estimation is about 500 msec for a 300 dpi document but the skew angle estimation takes only about 80 msec. The corresponding times for a 600 dpi document are 1.3 secs and 120 msec.

For most applications the skew angle is less than 10 degrees in either portrait or landscape. The algorithm successfully corrects the skew angle for angles in the range of ±30 degrees with accuracy higher than 0.02 degrees. For the given document [figure 2] skew angle was found to be -2.912 degrees. It must be noted that the first object, the long black line which is a defect from the photocopy machine, was estimated to have a skew angle of -3.036 degrees and when it was measured in an image editor was found -3.030 degrees. The second object (the title) was estimated to have a skew angle of -2.832 degrees and it was measured in an image editor to be -2.840 degrees. The error in the estimation skew angle for the specific objects is less than 0.01 degrees.

In the document shown in [figure 5] text and images are mixed. The skew angle was estimated to be –4.90 degrees. The above result is very accurate, even if the document contains images. The document was rotated by -5 degrees from its initial horizontal position. It must be noted that the orientation of the logo on the top left of the page does not follow the page orientation because of its abnormal shape. This object was not included in the skew angle estimation since its orientation exceeds the predefined limit from the mean skew angle.



**The JPEG Still Picture Compression Standard**

Gregory K. Wallace
Multimedia Engineering
Digital Equipment Corporation
Maynard, Massachusetts

*Submitted in December 1991 for publication in IEEE Transactions on Consumer Electronics*

*This paper is a revised version of an article by the same title and author which appeared in the April 1991 issue of* Communications of the ACM.

**Abstract**

For the past few years, a joint ISO/CCITT committee known as JPEG (Joint Photographic Experts Group) has been working to establish the first international compression standard for continuous-tone still images, both grayscale and color. JPEG's proposed standard aims to be generic, to support a wide variety of applications for continuous-tone images. To meet the differing needs of many applications, the JPEG standard includes two basic compression methods, each with various modes of operation. A DCT-based method is specified for "lossy" compression, and a predictive method for "lossless" compression. JPEG features a simple lossy technique known as the Baseline method, a subset of the other DCT-based modes of operation. The Baseline method has been by far the most widely implemented JPEG method to date, and is sufficient in its own right for a large number of applications. This article provides an overview of the JPEG standard, and focuses in detail on the Baseline method.

**1 Introduction**

Advances over the past decade in many aspects of digital technology - especially devices for image acquisition, data storage, and bitmapped printing and display - have brought about many applications of digital imaging. However, these applications tend to be specialized due to their relatively high cost. With the possible exception of facsimile, digital images are not commonplace in general-purpose computing systems the way text and geometric graphics are. The majority of modern business and consumer usage of photographs and other types of images takes place through more traditional analog means.

The key obstacle for many applications is the vast amount of data required to represent a digital image directly. A digitized version of a single, color picture at TV resolution contains on the order of one million bytes; 35mm resolution requires ten times that amount. Use of digital images often is not viable due to high storage or transmission costs, even when image capture and display devices are quite affordable.

Modern image compression technology offers a possible solution. State-of-the-art techniques can compress typical images from 1/10 to 1/50 their uncompressed size without visibly affecting image quality. But compression technology alone is not sufficient. For digital image applications involving storage or transmission to become widespread in today's marketplace, a standard image compression method is needed to enable interoperability of equipment from different manufacturers. The CCITT recommendation for today's ubiquitous Group 3 fax machines [17] is a dramatic example of how a standard compression method can enable an important image application. The Group 3 method, however, deals with bilevel images only and does not address photographic image compression.

For the past few years, a standardization effort known by the acronym JPEG, for Joint Photographic Experts Group, has been working toward establishing the first international digital image compression standard for continuous-tone (multilevel) still images, both grayscale and color. The "joint" in JPEG refers to a collaboration between CCITT and ISO. JPEG convenes officially as the ISO committee designated JTC1/SC2/WG10, but operates in close informal collaboration with CCITT SGVIII. JPEG will be both an ISO Standard and a CCITT Recommendation. The text of both will be identical.

Photovideotex, desktop publishing, graphic arts, color facsimile, newspaper wirephoto transmission, medical imaging, and many other continuous-tone image applications require a compression standard in order to

**Figure 4.** The document after the skew angle correction.

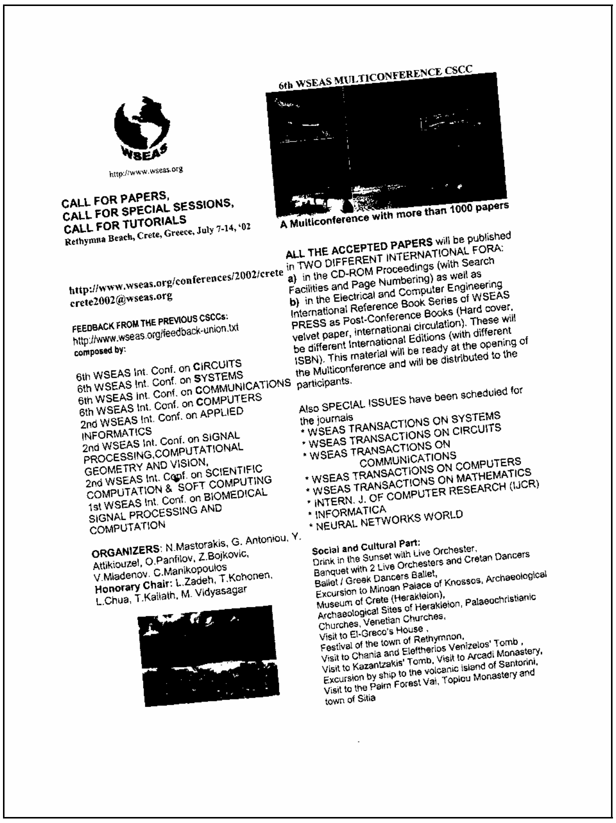
Proper assessment of algorithm performance requires measurement of the real skew angle of the document and comparison with the estimated value. In most cases this is very difficult because every line may have a skew angle that is slightly different from the others. One method to measure the accuracy of the algorithm is described below.

a) We consider an image and we estimate the skew angle $s_1$ which corresponds to a true angle of $r_1$ degrees. The true angle can not be exactly measured.
b) We rotate the image using an angle of $r_2$ degrees and we estimate the skew angle of $s_2$ degrees.
c) We rotate again the image by $r_3$=-$s_2$ and estimate the skew angle $s_3$. This estimation must be the same with $s_1$, thus $s_3$=$s_1$ if the document in step (a) is perfectly horizontal and the estimation introduces no errors.
d) We rotate again the document by $r_4$=-$r_3$ and we estimate the skew angle of $s_4$ degrees. The

document is in the same angle it was in step (b). The error between the two estimated skew angles is caused primarily by the rotation process.

In the document shown in [figure 4] we get the following values:

a) $r_1=0$ and $s_1=0.046$
b) $r_2=-3$ and $s_2=-2.912$
c) $r_3=-s_2=2.91$, $s_3=0.035$ and $e_1=s_3-s_1=0.011$
d) $r_4=-r_2=-2.91$, $s_4=2.894$ and $e_2=s_4-s_2=0.018$



WSEAS
http://www.wseas.org

CALL FOR PAPERS,
CALL FOR SPECIAL SESSIONS,
CALL FOR TUTORIALS
Rethymna Beach, Crete, Greece, July 7-14, '02

http://www.wseas.org/conferences/2002/crete
crete2002@wseas.org

FEEDBACK FROM THE PREVIOUS CSCCs:
http://www.wseas.org/feedback-union.txt
composed by:

6th WSEAS Int. Conf. on CIRCUITS
6th WSEAS Int. Conf. on SYSTEMS
6th WSEAS Int. Conf. on COMMUNICATIONS
6th WSEAS Int. Conf. on COMPUTERS
2nd WSEAS Int. Conf. on APPLIED INFORMATICS
2nd WSEAS Int. Conf. on SIGNAL PROCESSING,COMPUTATIONAL GEOMETRY AND VISION,
2nd WSEAS Int. Conf. on SCIENTIFIC COMPUTATION & SOFT COMPUTING
1st WSEAS Int. Conf. on BIOMEDICAL SIGNAL PROCESSING AND COMPUTATION

ORGANIZERS: N.Mastorakis, G. Antoniou, Y. Attikiouzel, O.Panfilov, Z.Bojkovic, V.Mladenov. C.Manikopoulos
Honorary Chair: L.Zadeh, T.Kohonen, L.Chua, T.Kailath, M. Vidyasagar

6th WSEAS MULTICONFERENCE CSCC

A Multiconference with more than 1000 papers

ALL THE ACCEPTED PAPERS will be published in TWO DIFFERENT INTERNATIONAL FORA:
a) in the CD-ROM Proceedings (with Search Facilities and Page Numbering) as well as
b) in the Electrical and Computer Engineering International Reference Book Series of WSEAS PRESS as Post-Conference Books (Hard cover, velvet paper, international circulation). These will be different International Editions (with different ISBN). This material will be ready at the opening of the Multiconference and will be distributed to the participants.

Also SPECIAL ISSUES have been scheduled for the journals
* WSEAS TRANSACTIONS ON SYSTEMS
* WSEAS TRANSACTIONS ON CIRCUITS
* WSEAS TRANSACTIONS ON COMMUNICATIONS
* WSEAS TRANSACTIONS ON COMPUTERS
* WSEAS TRANSACTIONS ON MATHEMATICS
* INTERN. J. OF COMPUTER RESEARCH (IJCR)
* INFORMATICA
* NEURAL NETWORKS WORLD

Social and Cultural Part:
Drink in the Sunset with Live Orchester,
Banquet with 2 Live Orchesters and Cretan Dancers
Ballet / Greek Dancers Ballet,
Excursion to Minoan Palace of Knossos, Archaeological Museum of Crete (Herakleion),
Archaeological Sites of Herakleion, Palaeochristianic Churches, Venetian Churches,
Visit to El-Greco's House ,
Festival of the town of Rethymnon,
Visit to Chania and Eleftherios Venizelos' Tomb ,
Visit to Kazantzakis' Tomb, Visit to Arcadi Monastery,
Excursion by ship to the volcanic island of Santorini,
Visit to the Palm Forest Vai, Topiou Monastery and town of Sitia

**Figure 5.** Document containing images and text.

The error $e_1$ indicates a possibly not totally horizontal document and the error $e_2$ degrees indicates that the rotation of the document introduces a small error since $r_4-r_3=0$. If rotation is ideal the error $e_2$ must be 0 but in practice is about 0.018. In conclusion we can not estimate the error introduced by our method exactly but the experiments show that the error is certainly less than 0.02 degrees.

## 5. CONCLUSION

In this paper an efficient, high precision algorithm is presented for the correction of the skew angle of scanned document images. The proposed method works in any kind of documents even when they contain images mixed with text. The skew angle can be corrected in every separate text line or in the whole document. The only limitation is that the skew angle of the document must be in the range of ±30 degrees. Documents' exceeding this limit is a very rare case. The proposed method is fast and can be used in demanding OCR applications. Finally a calibration method is described for the measurement of the accuracy of the skew angle estimation.

## REFERENCES


[1] Tzong-Wann Kao, Shi-Jinn Horng, Yue-Li Wang and Kuo-Liang Chung, A Constant Time Algorithm for Computing Hough Transform, *Pattern Recognition, Vol. 26, No. 2*, 1993, 277-286.

[2] S. C. Hinds, J. L. Fisher and D. P. D'Amato, A document skew detection method using run-length encoding and the Hough transform, *10th International Conference on Pattern Recognition 1*, 1990, 464-468.

[3] Hong Yan, Skew correction of document images using interline cross correlation, *CVGIP, Vol. 55, No. 6*, November 1993, 538-543.

[4] Teruo Akiyama and Norihiro Hagita, Automated Entry System for Printed Documents, *Pattern Recognition, Vol. 23, No. 11*, 1990, 1141-1154.

[5] H.S. Baird, The skew angle of printed documents, *in Proc. SPSE 40th Conf. Symp. Hybrid Imaging Systems*, Rochester NY, May 1987, 21-24.

[6] T. Saitoh and T. Pavlidis, Page segmentation without rectangle assumption, *in Proc. of 10th Int. Conf. on Pattern Recognition*, 1992, 277-280.

[7] Daniel S. Le, George R. Thoma and Harry Wechsler, Automated page orientation and skew angle detection for binary document images, *Pattern recognition, Vol. 10, No. 10*, 1994, 1325-1344.

[8] I. Pitas, *Digital image processing algorithms,* Prentice Hall, 1992.

[9] Henry S. Baird and Ken Thompson, Reading Chess, *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12,* No. 6, June 1990.

[10] N. Liolios, N. Fakotakis, G. Kokkinakis, On The Generalization of the Form Identification and Skew Detection Problem, *Pattern Recognition*, Vol. 35 , 2002, 253-264.

[11] G. Kapogiannopoulos, N. Koutsoulis, K. Koutroumbas, N. Kalouptsidis, A Novel Handwritten Character Recognition System, *Pattern Recognition*, under review.