Cohen's Class Distributions for Skew Angle Estimation in Noisy Ancient Arabic Documents

Nazih Ouwayed, Abdel Belaïd, François Auger

HAL Id: inria-00435253
https://hal.inria.fr/inria-00435253

Submitted on 24 Nov 2009

Cohen's Class Distributions for Skew Angle Estimation in Noisy Ancient Arabic Documents

N. Ouwayed
LORIA-University of Nancy 2
Campus Scientifique B.P. 239
54506 Vandoeuvre-Lès-Nancy
nazih.ouwayed@loria.fr

A. Belaïd
LORIA-University of Nancy 2
Campus Scientifique B.P. 239
54506 Vandoeuvre-Lès-Nancy
abelaid@loria.fr

F. Auger
University of Nantes
IREENA
44600 Saint-Nazaire
francois.auger@univ-nantes.fr

ABSTRACT
This paper presents an adaptive method for the skew angle estimation of noisy handwritten Arabic documents using the energy distributions of Cohen's class. The presence of noise in raw document images can create local maxima which disturb the projection histogram analysis. To avoid this drawback, we propose to use Cohen's class distributions on the projection histogram profiles obtained using different projection angles. These distributions reduce the importance of these local maxima and are fitted to the non-stationary nature of these histogram profiles. The orientation of the document is given by the highest maximum. The proposed skew angle detection technique has been evaluated on 864 skewed documents. The results of each distribution are presented at the end of this paper.

1. INTRODUCTION
Ancient Arabic textual archives contain a heavy volume of handwritten documents that need to be scanned and indexed. Some of these documents can be skewed which may handicap their recognition. Figure 1 shows some examples stemmed from the database studied [1, 2, 3]. We are looking for a method that can detect this orientation, whatever the size of the document.
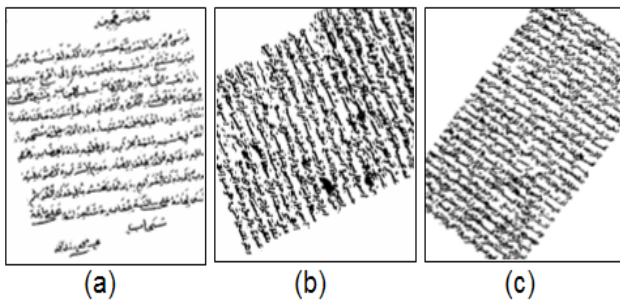

Figure 1: Samples of skewed Arabic handwritten documents.

The literature proposes some solutions for image document skew angle estimation. The projection techniques [12] seem the most appropriate ones but need to be adapted to Arabic documents. In fact, in Arabic, the words are composed of PAWs (Parts of Arabic Words) which are almost vertical or oblique, distorting the line orientation calculation. Furthermore, the gaps between PAWs and words are very narrow, essentially in the case of handwritten texts (see figure 2.b). This prevents to apply techniques like nearest neighbors as in [11] because of the alignment absence, or techniques based on the Hough Transform [7] because of the difficulty of locating voting points in the box bottoms. Sharif et al. [14] proposed two projection techniques for printed Persian and Arabic documents, based on black and white line strips. As these methods fail in large skew angles, they propose a slope seeker algorithm based on the Hough transform. Sarfraz et al. [13] used the Haar wavelet to decompose the image into sub-images at various level. In each level, they apply Principal Component Analysis(PCA) to estimate the orientation of the main axis (i.e. horizontal, vertical and diagonal). Although these techniques fit well to printed documents, they remain inadequate to handwritten documents, in which the interline distance is random and the skew angle can be large. E. Kavallieratou et al. [9] employed Cohen's class distributions on Latin documents. But no explanation was given on the relationship between the distributions properties and the results. We propose in this article to explore these properties and to find, thanks to them, the best distribution to use.

This paper is organized as follows. In section 2, we present the projection histogram profile analysis and the orientation estimation principle. In section 3, the Cohen's class is briefly presented, considering its distributions and their properties. In section 4, we show the application of the Cohen's distributions to the skew angle estimation. We present some experimental results in section 5 and a conclusion and future trends of this work in section 6.

2. ORIENTATION ESTIMATION
The idea of the usual projection technique is to find the projection angle for which the number of peaks of the histogram profile is as close as possible to the number of lines. The document orientation will then be equal to that angle. As the line number is usually unknown, we traditionally base the projection analysis on the peak sharpness and the valley depth. The orientation angle corresponds to the projection that gives the maximum difference between peaks and valleys [4, 12] (see figure 2.a with an angle equal to ($0°$) and where the high difference between peaks and valleys is representative to the three horizontal lines). But this method does not always yield the correct skew estimation because of line distortions and variabilities, as is usually the case in ancient Arabic documents.
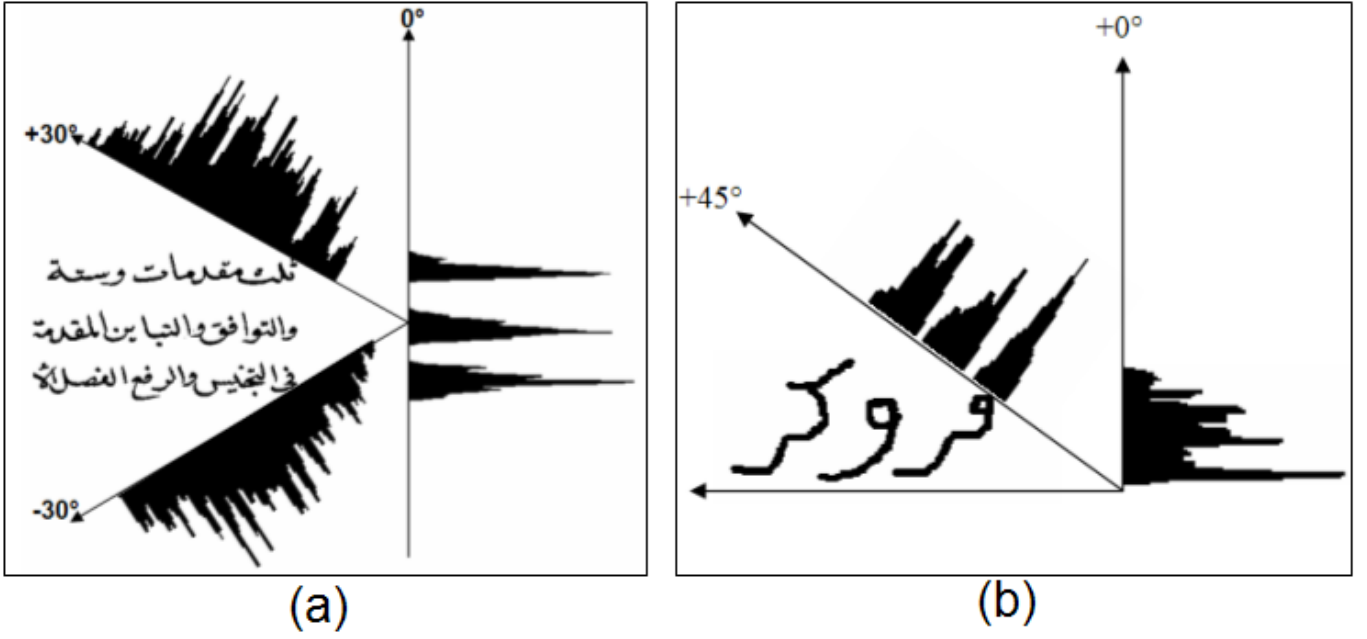
Figure 2: Depth and sharpness of the histogram profile computed for projection angles of $-30\,^{\circ}$, $0\,^{\circ}$ and $30\,^{\circ}$ and for a document made of three handwritten Arabic lines(a); Orientation error provided by local PAWs in an Arabic horizontal line (b).

3.   TIME-FREQUENCY DISTRIBUTIONS

Traditional signal processing approaches of such as the Fourier transform cannot describe the signal variation over both time and frequency. The energetic time-frequency distributions go beyond by analyzing the "non-stationarity" of a signal and distribute the energy of a signal in a time and frequency plane.

According to [6], the energy E_x of a signal $x(t)$ has the following expression:

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt = \int_{-\infty}^{+\infty} |\widehat{x}(f)|^2 \, df \qquad (1)$$

where $\widehat{x}(f)$ is the Fourier transform of the signal $x(t)$. The value E_x is quadratic. Therefore, the time-frequency distributions should keep this property.

In 1966, Cohen [6, 5] proved that a significant number of time-frequency distributions can be seen as particular cases of the following general expression:

$$C_x(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi_{r-d}(\tau,\xi)A_x(\tau,\xi)e^{j2\pi(t\xi - f\tau)}d\xi d\tau, \quad (2)$$

where $A_x(\tau,\xi)$ is the ambiguity function defined by:

$$A_x(\tau,\xi) = \int_{-\infty}^{+\infty} x(t+\tau/2)x^*(t-\tau/2)e^{-j2\pi\xi t}dt$$

The Cohen's class contains all the time-frequency distributions that are covariant under time- and frequency-shifts. The members of this class are identified by a particular kernel $\phi_{dD}(\tau,\xi)$, (expressed here in the delay-Doppler plane dD) which determines their theoretical properties and their practical readability. The properties

related to the skew angle estimation problem are summarized in Table 1. For the problem considered here, we think that properties P1, P2, P3 and P4 are highly necessary, whereas properties P5, P6 and P7 may help to obtain accurate estimations.

4.   APPLICATION OF THE COHEN'S DISTRIBUTIONS TO THE SKEW ANGLE ESTIMATION

4.1   Justification of the proposed method

When computing the histogram profile of a document along one direction of projection, we obtain, if this direction is the real orientation of the document, a histogram in which each line leads to a clearly localized local maximum (see figure 2). Each block of the document leads in the projection histogram to a succession of periodic peaks and valleys, whose period is relatively constant. This periodic succession is delimited by the block size ("time" support) and oscillates at a frequency determined by the space width between the lines.

As all the pixels are accumulated in the same positions, the local maxima have higher energy levels than with other projection directions. This explains why we can estimate the orientation of a document by seeking the projection angle for which the time-frequency distribution localizes a large energy level on a small area of the time-frequency plane. For example, Fig. 4 shows the increase of the maximum of the Wigner-Ville distribution when the number of peaks and valleys increases and when the valleys become wider.

4.2   Implementation Details

To estimate the orientation angle, we use the analytic signal $x_a(t)$ of the centered projection profile $x(t)$ of the document or of the selected area. The analytic signal is the signal $x(t)$ without its negative frequencies. The profile $x(t)$ is determined by projecting

	Property	Formula	Explanation
P1	Real-valued	$\forall t, f, C_x(t,f) = C_x^*(t,f)$, where $C_x^*(t,f)$ is the conjugate of $C_x(t,f)$	Allows to associate a real number to each point of the time-frequency plane and not a complex number: this reduces the amount of memory for the storage of the distribution and to find the highest value of a distribution (see figures 3.a and 3.b)
P2	Translation invariance	$\forall t_1, f_1, y(t) = x(t - t_1)e^{j2\pi f_1 t} \Rightarrow C_y(t,f) = C_x(t - t_1, f - f_1)$	A simple shift of the text in the image will result in a shift of the time-frequency representation: this shows that the distribution is not sensitive to the text location in the document (see figure 3.b and 3.c)
P3	Scale change co-variance	$\forall a > 0, y(t) = a^{-1/2}x(t/a) \Rightarrow C_y(t,f) = C_x(t/a, af)$	An expansion or reduction of the document will result in an expansion or compression of his projection profile: this will result in a correct change of its energy distribution (see figure 3.b)
P4	Energy conservation	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C_x(t,f)dtdf = \int_{-\infty}^{+\infty} \|x(t)\|^2 \, dt$	The energy of the projection profile is exactly distributed in the time-frequency plane
P5	Positivity	$\forall x(t), \forall t, f, C_x(t,f) \geq 0$	Enables us to more easily interpret the values of representation as energy measurements
P6	Signal null value conservation	$x(t_0) = 0 \Rightarrow \forall f, C_x(t_0, f) = 0$	Allows to get a null representation at all the "frequencies" at "instants" where the profile projection is zero: this correctly localizes the valleys in the projection profile (see figure 3.d)
P7	Signal support conservation	if $\forall t, \|t\| > T \Rightarrow x(t) = 0$ then $\forall t, \|t\| > T \Rightarrow C_x(t,f) = 0$	Allows to get a null time-frequency representation outside the range within which the projection profile has non-zero values

Table 1: Properties that a time-frequency distribution can satisfy and which are interesting for the skew angle estimation problem.

Distribution	Kernel	P1	P2	P3	P4	P5	P6	P7
Wigner-Ville	1	x	x	x	x			x
Pseudo Wigner-Ville	$h(\tau)$	x	x		x			x
Smoothed Pseudo Wigner-Ville	$h(\tau)\widehat{g}(\xi)$	x	x		x			
Margenau-Hill	$\cos(\pi\xi\tau)$	x	x	x	x		x	x
Pseudo Margenau-Hill	$h(\tau)\cos(\pi\xi\tau)$	x	x		x		x	x
Spectogram	$A_h^*(\tau, \xi)$	x	x		x	x		
Choi-Williams	$\exp(-4\pi^2\xi^2\tau^2/\sigma)$	x	x	x	x			
Born-Jordan	$\mathrm{sinc}(\pi\xi\tau)$	x	x	x	x			x
Zhao-Atlas-Marks	$h(\tau)\|\tau\|\mathrm{sinc}(\pi\xi\tau)$	x	x					x

Table 2: Some time-frequency distributions of the Cohen's class, their kernels and their properties.
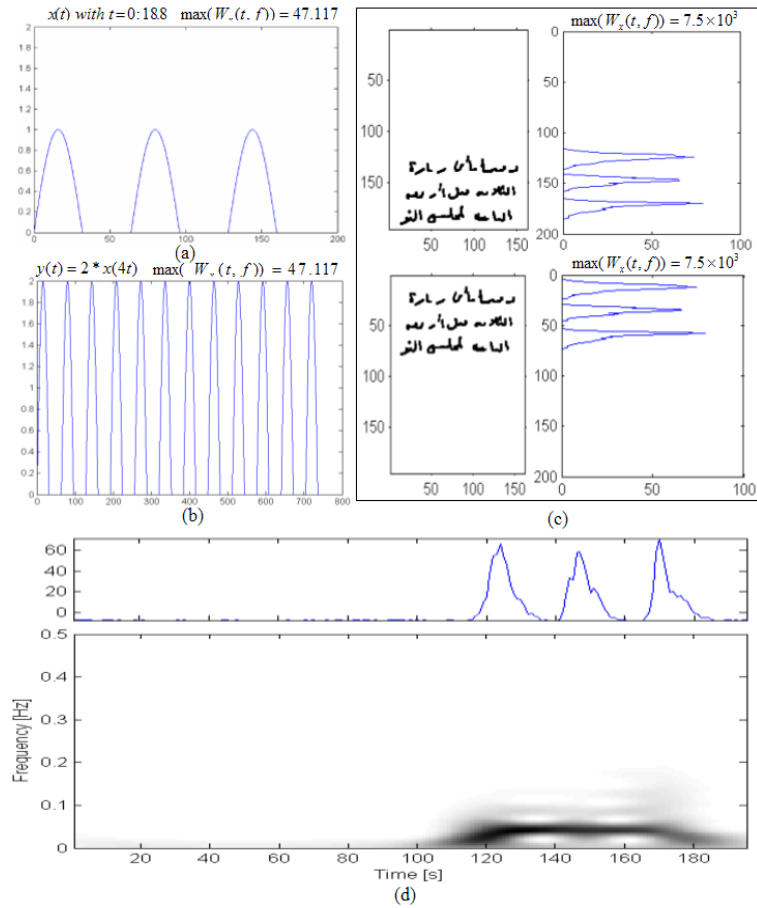
Figure 3: This figure showing on some projection histogram profiles more or less sharp, the property of the Cohen's class distributions and especially those of the Wigner-Ville ((a),(b),(c)) and Margenau-Hill (d) (the signal $x(t)$ used in (a) and (b) is equal to $\max(\sin(2\pi\frac{t}{64}), 0)$.
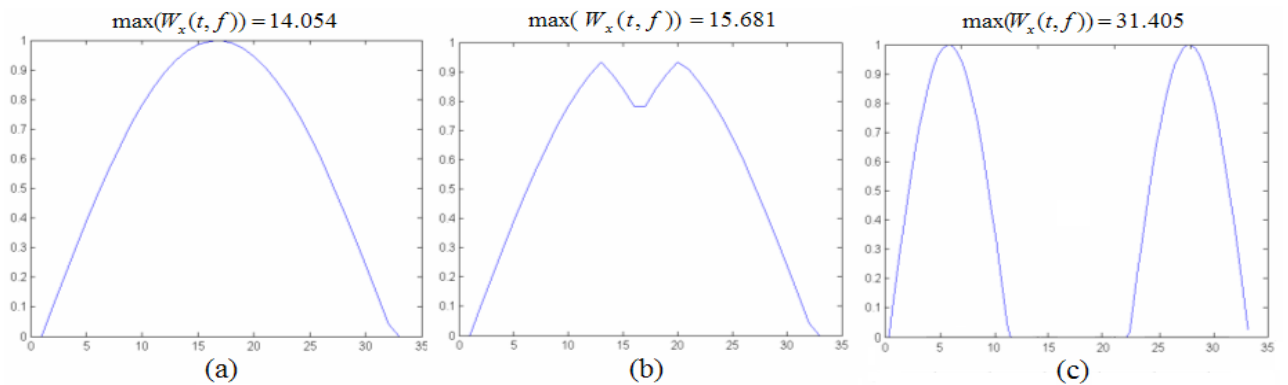

Figure 4: Examples of maximum value of the Wigner-Ville distribution obtained for different projection profiles.
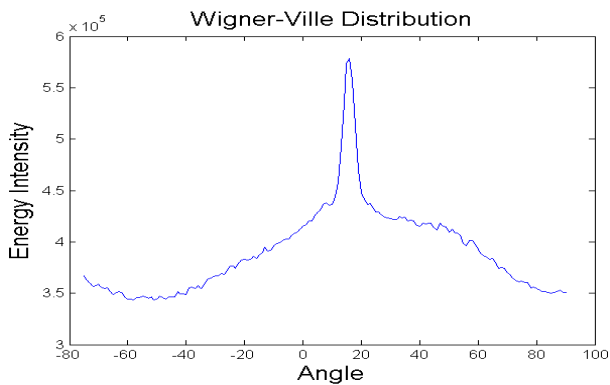
Figure 5: Intensity energy chart giving the Wigner-Ville distribution of the projection histogram profile of the document of the figure 1.a, by varying the projection angle by $+1°$.

Distribution	Estimation rate (%)	Average time (s)	Estimated Angle
Wigner-Ville	100	4.2	$+15.4°$
Pseudo Wigner-Ville	90	7.1	$+13.1°$
Smoothed Pseudo Wigner-Ville	96	6.9	$+14.3°$
Margenau-Hill	95	7.2	$+13.7°$
Pseudo Margenau-Hill	98	8.1	$+15.1°$
Spectrogram	90	7.9	$+12.9°$
Choi-Williams	97	7.6	$+15.2°$
Born-Jordan	87	7.8	$+11.8°$
Zhao-Atlas-Marks	92	8.5	$+12.6°$
Projection profile	93	6.4	$+14.9°$

Table 3: For a chosen set of Cohen's class distributions, estimated skew orientation angle for the document shown on Fig. 1.a, average computation time and average estimation rate obtained with 864 skewed documents.

each document with a chosen orientation. To calculate all possible profiles, we turn the image around its center of gravity (which gives us a point deduced from the image content and not from its size and framing) and we choose the horizontal axis as an arbitray reference for the zero degree angle. Then, we compute a time-frequency representation for each projection profile. The angle corresponding to the profile with the highest maximum value of its time-frequency representation is chosen as the estimated angle of the document. Fig. 5 shows the highest values of the Wigner-Ville distribution (WVD) of the histogram profiles of the document shown on Fig. 1.a. The third column of Table 3 shows the skew angle of the document shown on Fig. 1.a. estimated with some Cohen's class distributions. The correct value is $15.4°$.

5. EXPERIMENTAL RESULTS AND DISCUSSION

To study the effectiveness of our approach, we have experimented it on 864 Arabic handwritten documents. These documents have different sizes, contain several types of writing, layout (1 column, 2 columns), a mix of text and tables, etc. The experiments were prepared after a manual orientation of the documents into different angles ranging from $-75°$ to $+90°$. The tests were carried out on a PC equipped with a microprocessor Pentium M 1.4 GHz

and 1 GB of memory under Windows XP. The execution time is measured in seconds. It depends on the size of each document and the distribution. Table 3 (third column), shows that the precision obtained is very high, varying around $0.5°$.

As shown on Table 3, the Wigner-Ville distribution reaches the highest estimation rate (100%). The other distributions yield a lower estimation rate, either because they do not satisfy properties that are important for the skew angle estimation problem, such as the scale invariance property and the support conservation, because their localization of the signal components is not sufficiently precise to provide a skew angle estimation with the maximum of the representation, or because the parameters of these distributions are not fitted the the analysed histogram profiles. These parameters are

- the time smoothing window $g(t)$;
- the frequency smoothing window $h(\tau)$;
- the width of the Choï-Williams kernel distribution σ.

To simplify the comparisons, all smoothing windows have been chosen of Kaiser-Bessel type [8]. The best results have been obtained when the length of the time smoothing window

and the length of the frequency smoothing

window are constant and respectively equal to 19 and 91 time samples. Table 3 also shows the results obtained with the projection profile analysis method recalled in section 2, which is classically used for handwritten documents. The skew angle estimated with the Wigner-Ville distribution is much better than the projection method one and moreover, it is faster.

6. CONCLUSION AND FUTURE WORK

We have proposed in this article an original approach which seeks to estimate the orientation of an Arabic document. Because of the local variation of the handwritten lines and because of the vertical and slant ligatures in the Arabic script, the classical methods based on nearest neighbours or on the Hough transform often fail. Hence, we used the time-frequency distributions of the projection profile for the angle determination. We first studied the Cohen's class distributions and examined their properties. We have chosen the Wigner-Ville distribution whose properties fit to the document analysis requirements. This distribution allows us to precisely determine the orientation angle by maximizing the signal energy. Unlike the other distributions, the Wigner-Ville distribution verifies two interesting properties: scale invariance and time support conservation. This method has been experimented on 864 skewed handwritten Arabic documents. The estimation rate varies depending on the characteristics of each distribution. The Wigner-Ville distribution gives the highest estimation rate which is equal to 100%. In a previous work [10], we used this approach for the analysis of multi-oriented Arabic documents. The distribution was applied on small windows and orientation zones were delimited by the local extension of similar orientation windows. We are going to improve this approach by testing it on more complex documents including different orientations. The next step will concern the effective line extraction exploiting the local orientation knowledge.

7. AKNOWLEGEMENTS

We wish to thank the libraries that provide these documents (Tunisian National Library [1], National Library of Medicine in the USA [2] and the National Library and Archives of Egypt [3]).

8. REFERENCES

[1] http://www.bibliotheque.nat.tn/.

[2] http://www.nlm.nih.gov/hmd/arabic/welcome.html.

[3] http://portal.unesco.org/ci/photos/showgallery.php/cat/559.

[4] T. Akiyama and N. Hagita. Automatic entry system for printed documents. *Pattern Recognition*, 23:1141–1154, 1990.

[5] L. Cohen. Generalized phase-space distribution functions. *J. Math. Phys.*, 7(5):781–786, May 1966.

[6] F. Hlawatsch and F. Auger. *Time-frequency analysis: concepts and methods*. ISTE-John Wiley and sons, 2008.

[7] H. F. Jiang, C. C. Han, and K. C. Fan. A fast approach to the detection and correction of skew documents. *Pattern Recognition Letters*, 18:675–686, 1997.

[8] J.F. Kaiser and R.W. Schafer. On the use of the io-sinh window for spectrum analysis. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-28(1):105–107, 1980.

[9] E. Kavallieratou, N. Fakotakis, and G. Kokkinakis. Skew angle estimation in document processing using cohen's class distributions. *Pattern Recogn. Lett.*, 20:11–13, 1999.

[10] N. Ouwayed and A. Belaid. Multi-oriented text line extraction from handwritten arabic documents. *The Eighth IAPR Workshop on Document Analysis Systems (DAS 2008)*, pages 339–346, 2008.

[11] U. Pal and B. B. Chaudhuri. An improved document skew angle estimation technique. *Pattern Recognition Letters*, 17:899–904, 1996.

[12] T. Pavlidis and J. Zhou. Page segmentation and classification. *Computer Vision Graphics and Image Processing*, 54(2):484–496, 1992.

[13] M. Sarfraz, A. Zidouri, and S. A. Shahab. A novel approach for skew estimation of document images in ocr system. In *CGIV '05: Proceedings of the International Conference on Computer Graphics, Imaging and Visualization*, pages 175–180, Washington, DC, USA, 2005. IEEE Computer Society.

[14] A. E. Sharif and N. Movahhedinia. On skew estimation of persian/arabic printed documents. *Journal of Applied Sciences*, 8(12):175–180, 2008.