2011 International Conference on Document Analysis and Recognition

Adaptive Zoning Features for Character and Word Recognition

B. Gatos[1], A. L. Kesidis[1,2] and A. Papandreou[1]

[1]Computational Intelligence Laboratory
Institute of Informatics and Telecommunications
National Center for Scientific Research "Demokritos"
GR-15310 Athens, Greece
{bgat, akesidis, alexpap}@iit.demokritos.gr

[2]Department of Surveying Engineering
Technological Educational Institution of Athens
GR-12210 Athens, Greece
akesidis@teiath.gr

Abstract—**Zoning features are of the most popular and efficient statistical features that provide high speed and low complexity for character and word recognition. They are calculated by the density of pixels or pattern characteristics in several zones we divide the pattern frame. In this paper, we introduce the idea of adaptive zoning features that are extracted after adjusting the position of every zone based on local pattern information. This adjustment is performed by moving every zone towards the pattern body. This process is based on the maximization of the local pixel density around each zone. We have extensively tested our approach for character and word recognition as well as for using the pixel density or pattern characteristics in every zone. For all cases, we have recorded a significant improvement when the zoning features are used in the proposed adaptive way.**

Zoning Features, Character Recognition, Word Recognition

I. INTRODUCTION

In character and word recognition systems, feature extraction is an important task in the recognition pipeline. It aims at extracting representative information from characters or words in order to minimize the within class pattern variability while enhancing the between class pattern variability. Good features make the subsequent classification task less difficult.

Feature extraction methods for character and word recognition have been based mainly on three types of features: a) statistical derived from statistical distribution of points, b) structural and c) transformation-based or moment-based features. The most common statistical features used for character and word representation are: a) zoning, where the image is divided into several zones [1-6], b) projections [7] and c) crossings and distances [8]. Structural features are based on topological and geometrical properties of the character/word, such as maxima and minima, reference lines, ascenders, descenders, cusps above and below a threshold, strokes and their direction between two points, horizontal curves at top or bottom, cross points, end points, branch points etc. [9]. Fourier Transform is often used to calculate transformation–based features [10]. The goal is to choose a magnitude spectrum of the measurement vector as the features in an n–dimensional Euclidean space. Moment-based features, such as Legendre or Zernike moments, form a compact representation of the original character that makes the process of recognition invariant to scale, transformation and rotation [11, 12]. A survey on feature extraction methods can be found in [13]. Moreover, other approaches focus on measuring the similarity/dissimilarity between shapes by mapping one pattern onto another ([14], [15]).

Features based on zones are of the most popular and efficient statistical features and provide high computational speed and low complexity for character and word recognition. They are calculated by the density of pixels or pattern characteristics in several zones we divide the pattern frame. In particular, standard zoning methods are defined according to a $N\times M$ regular grid superimposed on the image body [1]. Recently, zoning features based on pixel density have been combined with word profiles in a hybrid scheme for handwritten word recognition [2] as well for word spotting in historical printed documents [3]. In [4], features based on distances and angles of the skeleton pixels in each zone are used. Both neural networks and fuzzy logic techniques are then used for recognition. The methodology presented in [5] is based on the direction of the contour of the character by computing histograms of chain codes in each zone. In [6], it is observed that when the contour curve is close to zone borders, small variations in the contour curve can lead to large variations in the extracted features. For this reason, zones with fuzzy borders are introduced. Features detected near the zone borders are given fuzzy membership values to two or four zones.

In this paper, we propose an improved version of zoning features by introducing the idea of adaptive zones. According to our approach, features are extracted after adjusting the position of every zone based on local pattern information. This adjustment is achieved by moving every zone towards the pattern body. The offset that is used for adjusting zone position is calculated by maximizing the local pixel density around the zone. In Section II, we describe the proposed adaptive zoning features. We have extensively tested our approach for character and word recognition as well as for using the pixel density or pattern characteristics in every zone. As it shown in Section III, for all cases we have recorded a significant improvement when the zoning features are used in the proposed adaptive way. Conclusions and future work plans are given in Section IV.

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.234

II. THE PROPOSED ADAPTIVE ZONING FEATURES

A. Size Normalization

For the case of character recognition, the character image is normalized to a WxH matrix under the condition that the aspect ratio is preserved. Exact positioning of the character in the matrix is achieved by placing the geometric center of the character in the center of the matrix. A 60x60 size-normalized character is presented in Fig. 1.

For the case of word recognition, the positioning of the word in the size-normalized matrix WxH is achieved by placing the baseline areas of the word in the center of the matrix. This is achieved by placing the upper word baseline at vertical offset H/3 and the lower word baseline at vertical offset 2*H/3. Baseline detection is accomplished using a horizontal projections based approach. A similar approach is also used in [16]. A word size normalization example is presented in Fig. 2.

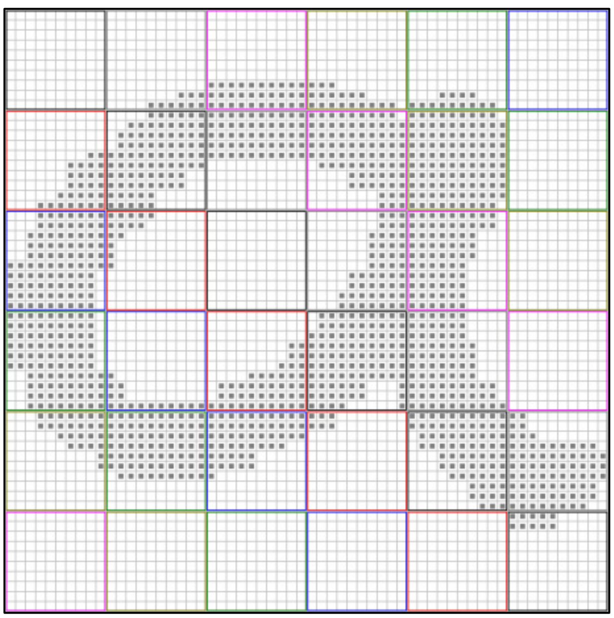

Figure 1. Zoning procedure example.

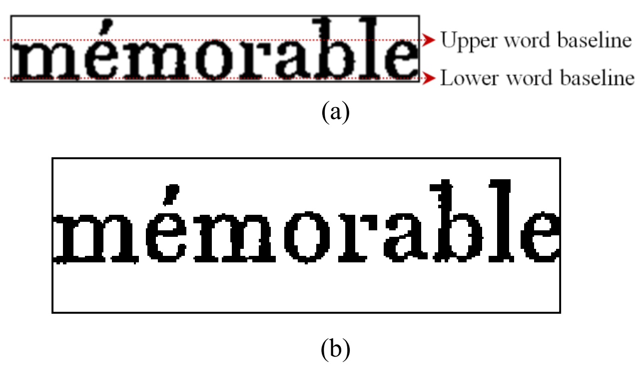

Figure 2. Word size normalization example: (a) orginal, (b) size-normalized image.

B. Zone Adjustment

Let suppose that a $N \times M$ regular grid is superimposed on the pattern image. If every grid window is of size KxΛ then the size-normalized pattern image $I(x,y) \in \{0,1\}$ has an overall size of WxH where $W=N*K$, $H=M*\Lambda$, x=0..W-1 and y=0..H-1. For example, in the standard zoning procedure of Fig.1 we have $N=M=6$, $K=\Lambda=10$, $W=H=60$.

In standard zoning methods, the coordinates of every zone (n,m) are defined as follows:

$$Z_{nm}^{x_1} = (n-1)K \tag{1}$$

$$Z_{nm}^{x_2} = nK - 1 \tag{2}$$

$$Z_{nm}^{y_1} = (m-1)\Lambda \tag{3}$$

$$Z_{nm}^{x_2} = m\Lambda - 1 \tag{4}$$

where n=1..N and m=1..M. In Fig. 3, the coordinates of zone (2,3) of Fig. 1 are presented. We have $Z_{23}^{x_1}$ = 10, $Z_{23}^{y_1}$ = 20, $Z_{23}^{x_2}$ = 19, $Z_{23}^{y_2}$ = 29.

We propose that the position of the zones can be adjusted based on local pattern information. More specifically, this adjustment is achieved by moving every zone towards pattern body. The offset that is used for adjusting zone position is calculated by maximizing the local pixel density around the zone as follows:

$$(dx_{nm}, dy_{nm}) = \arg\max_{x \in [-\lambda_x \ldots \lambda_x], y \in [-\lambda_y \ldots \lambda_y]} \sum_{i=z_{nm}^{x1}}^{z_{nm}^{x2}} \sum_{j=z_{nm}^{y1}}^{z_{nm}^{y2}} I(x+i, y+j) \tag{5}$$

Parameters λ_x and λ_y define the horizontal and vertical range for adjusting the position of the zones. Since large values for parameters λ_x and λ_y can affect the computational time needed for feature extraction, we propose that these parameters have to range between 1 and 3. An analysis of how these parameters affect the computational time is given in Section III.

The new coordinates of every zone (n,m) are defined as follows:

$$Z_{nm}^{x_1'} = Z_{nm}^{x_1} + dx_{nm} \tag{6}$$

$$Z_{nm}^{x_2'} = Z_{nm}^{x_2} + dx_{nm} \tag{7}$$

$$Z_{nm}^{y_1'} = Z_{nm}^{y_1} + dy_{nm} \tag{8}$$

$$Z_{nm}^{y_2'} = Z_{nm}^{y_2} + dy_{nm} \tag{9}$$

An example of adjusting the position of a zone is given in Fig. 3. In this example, zone (2,3) of Fig.1 is moved by offset (-2,-2) since according to eq. (5) we have calculated $\{dx_{23}, dy_{23}\}$ = {-2,-2}. Fig. 4 presents the result of the zone adjustment procedure for the image of Fig.1.

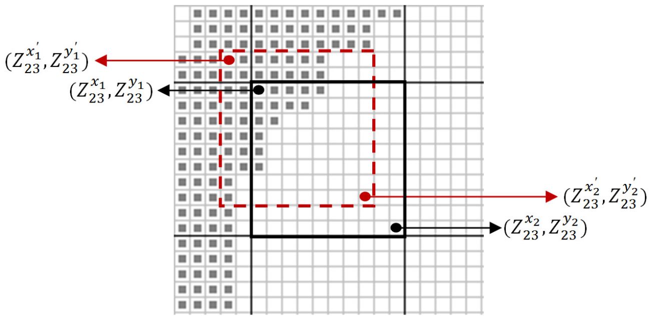

Figure 3. Zone position adjustment for zone (2,3) of Fig.1.

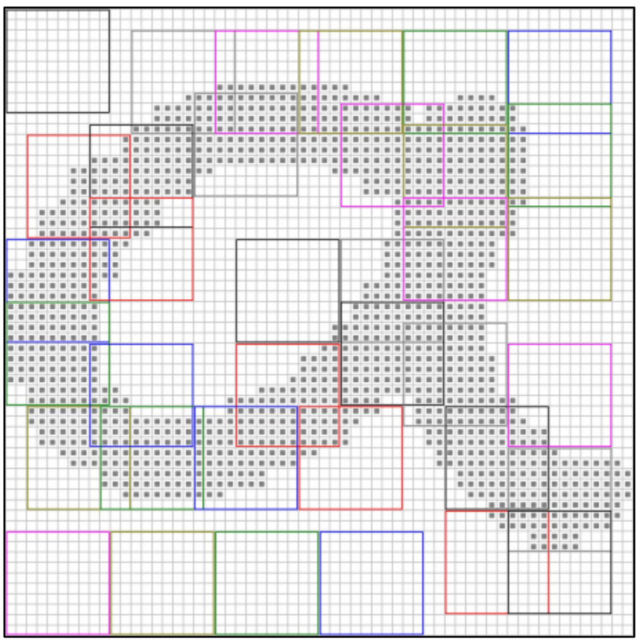

Figure 4. Position adjustment for all zones in the image of Fig.1.

C. Feature Extraction

We examine two types of zoning features: a) features that are calculated by the density of pixels and b) features that correspond to pattern characteristics in every zone.

Features based on pixel density are calculated directly from the size-normalized image I as follows:

$$denF_{nm} = \frac{1}{K * \Lambda} \sum_{x=Z_{nm}^{x_1'}}^{Z_{nm}^{x_2'}} \sum_{y=Z_{nm}^{y_1'}}^{Z_{nm}^{y_2'}} I(x, y) \tag{10}$$

In order to extract features based on zone characteristics we follow approach [4] according to which we first produce the skeleton of the size-normalized image. Then, we calculate the normalized distances of all skeleton pixels taking the top left corner of every zone as the absolute origin (0,0):

$$destF_{nm} = \frac{1}{P_{nm}\sqrt{K^2 + \Lambda^2}} \sum_{k=1}^{P_{nm}} d_k^{nm} \tag{11}$$

where P_{nm} is the number of skeleton pixels in zone (n,m) and d_k^{nm} is the distance of k skeleton pixel of zone (n,m) from point $(Z_{nm}^{x_1'}, Z_{nm}^{y_1'})$. Additionally, we calculate features based on the corresponding angles that are formed by the skeleton pixels:

$$angF_{nm} = \frac{1}{90 * P_{nm}} \sum_{k=1}^{P_{nm}} \theta_k^{nm} \tag{12}$$

where θ_k^{nm} = Arctan(-j/i) for a skeleton pixel k position (i,j). Again, the absolute origin (0,0) is the top left corner point $(Z_{nm}^{x_1'}, Z_{nm}^{y_1'})$ of every zone.

The total number of features based on pixel density is N*M while the total number of features based on zone characteristics is 2*N*M. All features range from 0 to 1.

III. Experimental Results

In this Section we present the results of testing the proposed adaptive pattern zoning for character and word recognition as well as for using the pixel density or pattern characteristics in every zone.

A. Character Recognition

For the character recognition experiments we used the CIL Database which consists of 28750 Greek handwritten characters (46 classes) [17]. The 4/5 of characters of each class was used for training (23000 characters) while the remaining 1/5 for testing (5750 characters). All characters were normalized to a 60x60 matrix (W=H=60) following the procedure described in Section IIA. Some samples of size-normalized characters of the CIL database are given in Fig. 5. For classification we used the Euclidean distance between two feature vectors combined with a minimum distance classifier. In Table I, we present the recognition accuracy achieved (a) for several number of zones N, M we divide the characters, (b) for several values for the parameters λ_x and λ_y that are used for adjusting the position of the zones, (c) for using features based on pixel density as well as on pattern characteristics as described in Section IIC. We applied the experiments for N=M={10,12,15} and λ_x=λ_y={0,1,2,3,4}. It should be noticed that the experiments with λ_x=λ_y=0 correspond to the standard zoning approach [1].

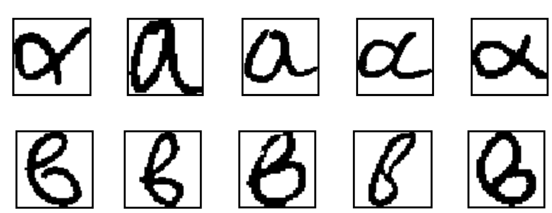

Figure 5. Character samples from the CIL database (letters “α” and “β”).

TABLE I. CHARACTER RECOGNITION RESULTS

		Standard Zones	Adaptive Zones			
		$\lambda_x=\lambda_y=0$	$\lambda_x=\lambda_y=1$	$\lambda_x=\lambda_y=2$	$\lambda_x=\lambda_y=3$	$\lambda_x=\lambda_y=4$
Pixel density features	N=M=10	85,98%	87,95%	**88,22%**	87,13%	85,46%
	N=M=12	85,48%	87,82%	**88,35%**	86,94%	85,41%
	N=M=15	84,61%	87,56%	**88,29%**	87,10%	85,52%
Features based on pattern characteristics [4]	N=M=10	66,80%	74,65%	76,73%	**77,08%**	75,34%
	N=M=12	61,47%	74,25%	78,05%	**78,76%**	76,21%
	N=M=15	52,32%	70,90%	73,76%	**73,85%**	71,14%

As it can be seen in Table I, in all cases the adaptive zones with $\lambda_x=\lambda_y=1..3$ outperform the standard zones ($\lambda_x=\lambda_y=0$). For the case of pixel density features, the best performance is achieved for $N=M=12$ and $\lambda_x=\lambda_y=2$ and corresponds to recognition accuracy of 88.35% while the best performance using the standard zones is 85.98%. For the case of features based on pattern characteristics [4], the best performance is achieved for $N=M=12$ and $\lambda_x=\lambda_y=3$ and corresponds to recognition accuracy of 78.76% while the best performance using the standard zones is 66.80%. In Table II we also present the time needed for training when different values for parameters λ_x, λ_y are used (no speed optimization is involved).

TABLE II. TIME NEEDED FOR TRAINING

	Standard Zones	Adaptive Zones			
	$\lambda_x=\lambda_y=0$	$\lambda_x=\lambda_y=1$	$\lambda_x=\lambda_y=2$	$\lambda_x=\lambda_y=3$	$\lambda_x=\lambda_y=4$
Time	18sec	24 sec	36 sec	53 sec	75 sec

B. Word Recognition

We tested our methodology for word recognition on a historical French book [18] that consists of 153 pages and 46197 words. The word segmentation as well as the ASCII ground truth were manually created. All words were normalized to a 300x90 matrix (W=300, H=90) following the procedure described in Section IIA. Every word is divided into 30x9 zones (N=30, M=9). Examples of size-normalized words from this set are presented in Fig. 6. We randomly selected five instances of the words 'France', 'Louis', 'famille', 'mort' and 'justice', thus yielding 25 queries in total. The total number of instances of those words is 44, 156, 47, 51 and 44, respectively. Let n_inst be the total number of instances of a word in the ground truth and n_corr the number of correct instances of the word in the first n_inst retrieved instances. The word retrieval performance can be calculated as follows:

$$Word\_Retrieval\_Performance = \frac{n\_corr}{n\_inst} \tag{13}$$

In Table III, we present the retrieval performance achieved (a) for several word queries, (b) for several numbers for the parameters λ_x and λ_y that are used for adjusting the position of the zones, (c) for using features based on pixel density as well as on pattern characteristics as described in Section IIC. We used again $\lambda_x=\lambda_y=\{0,1,2,3,4\}$ while experiments with $\lambda_x=\lambda_y=0$ correspond to using the standard zoning approach [1].

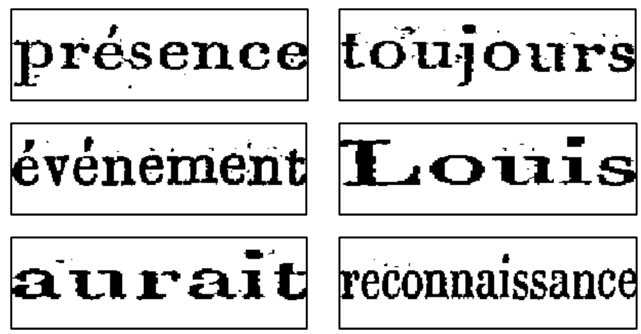

Figure 6. Sample of size-normalized words from our set.

TABLE III. WORD RETRIEVAL RESULTS

		Standard Zones	Adaptive Zones			
		$\lambda_x=\lambda_y=0$	$\lambda_x=\lambda_y=1$	$\lambda_x=\lambda_y=2$	$\lambda_x=\lambda_y=3$	$\lambda_x=\lambda_y=4$
Pixel density features	Query: "France"	92,73%	94,55%	95,00%	**95,91%**	95,00%
	Query: "Luis"	93,08%	94,10%	94,87%	**95,13%**	94,74%
	Query: "famille"	74,89%	75,74%	78,72%	81,70%	**83,83%**
	Query: "mort"	**85,10%**	83,53%	83,14%	81,57%	83,92%
	Query: "justice"	64,55%	67,73%	72,73%	77,73%	**81,82%**
	Total	85,67%	86,67%	88,07%	89,12%	**90,00%**
Features based on pattern characteristics [4]	Query: "France"	68,18%	80,00%	**82,73%**	75,00%	73,18%
	Query: "Luis"	83,21%	**91,15%**	88,08%	79,10%	84,10%
	Query: "famille"	65,11%	70,21%	**71,49%**	71,06%	63,40%
	Query: "mort"	**86,67%**	85,88%	83,53%	76,86%	50,20%
	Query: "justice"	57,73%	56,82%	58,64%	50,91%	**59,55%**
	Total	76,02%	**81,64%**	80,64%	73,51%	71,64%

As it can be seen in Table III, for the case of pixel density features, the best word retrieval performance is achieved for λ_x=λ_y=4 and corresponds to a total word retrieval performance of 90% while the performance using the standard zones is 85.1%. Significant improvement is also achieved for λ_x=λ_y=3 (word retrieval performance equals to 89.12%) as well as for λ_x=λ_y=2 (word retrieval performance equals to 88.07%). For the case of features based on pattern characteristics, the best performance is achieved for λ_x=λ_y=1 and corresponds to a total word retrieval performance of 81.64% while the performance using the standard zones is 76.02%.

IV. Conclusions and Future Work

In this paper, we introduced the idea of adaptive zoning features that are extracted after adjusting the position of every zone based on local pattern information. This adjustment is performed by moving every zone towards the pattern body. This process is based on the maximization of the local pixel density around each zone. We have extensively tested our approach for character and word recognition as well as for using the pixel density or pattern characteristics in every zone. For all cases, we have recorded a significant improvement when the zoning features are used in the proposed adaptive way. Concerning character recognition, the recognition accuracy improves from 85.98% to 88.35% when using pixel density features and from 66.80% to 78.76% when using pattern characteristics in each zone. Concerning word recognition, the retrieval accuracy improves from 85.1% to 89.12% when using pixel density features and from 76.02% to 81.64% when using several pattern characteristics in each zone. We have also recorded that we need 33% - 100% more time comparing to the standard zones in order to get an acceptable improvement.

Future work includes the investigation of a method in order to automatically select the best values for parameters λ_x and λ_y that define the horizontal and vertical range for adjusting the position of zones. Furthermore, we plan to work towards the acceleration of the adaptive zone procedure.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement n° 215064 (project IMPACT).

References

[1] M. Bokser , “Omnidocument Technologies”, Proc. IEEE, Vol. 80, pp. 1066-1078, 1992.

[2] B. Gatos, I. Pratikakis and S.J. Perantonis, "Hybrid Off-Line Cursive Handwritten Word Recognition", 18th International Conference on Pattern Recognition (ICPR2006), pp. 998-1001, Hong Kong, August 2006.

[3] T. Konidaris, B. Gatos, K. Ntzios, I. Pratikakis, S. Theodoridis and S. J. Perantonis, "Keyword-Guided Word Spotting in Historical Printed Documents Using Synthetic Data and User feedback", International Journal on Document Analysis and Recognition (IJDAR), special issue on historical documents, Vol. 9, No. 2-4, pp. 167-177, 2007.

[4] M. Hanmandlua, K.R. Murali Mohanb, S. Chakrabortyc, S. Goyald and D. Roy Choudhurye, “Unconstrained handwritten character recognition based on fuzzy logic”, Pattern Recognition, Vol. 36, pp. 603 – 623, 2003.

[5] K. M. Mohiuddin and J. Mao, "A Comprehensive Study of Different Classifiers for Hand-printed Character Recognition", Pattern Recognition, Practice IV, pp. 437- 448, 1994.

[6] V.L. Lajish, “Handwritten Character Recognition using Perceptual Fuzzy-Zoning and Class Modular Neural Networks”, Proc. 4th International Conference on Innovations in Information Technology, Innovations '07, DOI: 10.1109/IIT.2007.4430497, 2007.

[7] A. L. Koerich, "Unconstrained Handwritten Character Recognition Using Different Classification Strategies", Proc. International Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR), 2003.

[8] J. H. Kim, K. K. Kim, C. Y. Suen, "Hybrid Schemes Of Homogeneous and Heterogeneous Classifiers for Cursive Word Recognition", Proc. 7th International Workshop on Frontiers in Handwritten Recognition, Amsterdam, pp 433 – 442, 2000.

[9] N. Arica and F. Yarman-Vural, "An Overview of Character Recognition Focused on Off-line Handwriting", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 31, Issue 2, pp. 216 – 233, 2001.

[10] G. G. Rajput, S. M. Mali,”Fourier Descriptor based Isolated Marathi Handwritten Numeral Recognition”, Int’l Journal of Computer Applications, Volume 3 – No.4, doi: 10.5120/724-1017, June 2010.

[11] G. L. Cash and M. Hatamian. Optical Character Recognition by the Method of Moments, Computer Vision, Graphics, and Image Processing, Vol. 39, pp. 291-310, 1987.

[12] A. Khotanzd and Y. H. Hong, "Invariant Image Recognition by Zernike Moments", IEEE Trans. Pattern Anal. Machine Intell.,Vol. 12, Issue 5, pp. 489-497, 1990.

[13] O. D. Trier, A. K. Jain, T.Taxt, "Features Extraction Methods for Character Recognition – A Survey ", Pattern Recognition, Vol.29, No.4, pp. 641-662, 1996.

[14] S. Belongie, J. Malik, J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No. 4, pp. 509-522, 2002.

[15] A. K. Jain, D. Zongker, "Representation and Recognition of Handwritten Digits using Deformable Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 12, pp. 1386-1391, 1997.

[16] B. Gatos, I. Pratikakis, A.L. Kesidis and S.J. Perantonis, "Efficient Off-Line Cursive Handwritten Word Recognition", 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2006), pp. 121-125, La Baule, France, 2006.

[17] G. Vamvakas, B. Gatos, S. Petridis and N. Stamatopoulos, "An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek Handwritten Character Recognition", 9th International Conference on Document Analysis and Recognition (ICDAR'07), pp. 1073-1077, Curitiba, Brazil, 2007.

[18] Le Dernier fils de France, ou le Duc de Normandie, fils de Louis XVI et de Marie-Antoinette, par A. Bibliothèque nationale de France, 1838.