Writer Identification of Arabic Handwritten Digits

Sameh M. Awaida and Sabri A. Mahmoud
King Fahd University of Petroleum and Minerals, Dhahran 31261, KSA
{sameho, smasaad}@kfupm.edu.sa

Abstract

This paper addresses the identification of Arabic handwritten digits. In addition to digit identifiability, the paper presents digit recognition. The digit image is divided into grids based on the distribution of the black pixels in the image. Several types of features are extracted (viz. gradient, curvature, density, horizontal and vertical run lengths, stroke, and concavity features) from the grid segments. K-Nearest Neighbor and Nearest Mean classifiers are used. A database of 70000 of Arabic handwritten digit samples written by 700 writers is used in the analysis and experimentations.

The identifiability of isolated and combined digits are tested. The analysis of the results indicates that Arabic digits 3 (٣), 4 (٤), 8 (٨), and 9 (٩) are more identifiable than other digits while Arabic digit 0 (٠) and 1 (١) are the least identifiable. In addition, the paper shows that combining the writer's digits increases the discriminability power of Arabic handwritten digits. Combining the features of all digits, K-NN provided the best accuracy in text-independent writer identification with top-1 result of 88.14%, top-5 result of 94.81%, and top-10 results of 96.48%.

1. Introduction

Writer identification is the task of determining the author of sample handwriting from a set of writers. Writer verification is the process of comparing questioned handwriting with samples of handwriting obtained from known sources for the purposes of determining authorship or non-authorship *[1]*. The main applications of Arabic writer identification/verification systems are crime suspects' identification in forensic sciences, forgery detection, and in the identification of the writers of Arabic and Islamic manuscripts.

In classification problems such as writer, face, finger print, or speaker identification and verification, the number of classes is usually very large. We transform the 'many class' problem into a 'two class' problem by distinguishing between intra-class and inter-class distances, and noticing that in general intra-class variations are less than inter-class variations. In recent years, writer identification/verification has become a common application confirming the document authenticity in the financial district, or the identity of suspected criminals, etc. In May 13, 1999, the United States vs. Paul decided that Handwriting analysis qualifies as expert testimony and is therefore admissible *[2]*.

The literature review is addressed in Section 2; Section 3 presents a summarized description of the used features; the experimental results are detailed in Section 4; and finally, the conclusions are given in Section 5.

2. Current research and technology

Automatic, offline writer identification enjoys renewed interest [2-9]. The identification of a person on the basis of a handwritten sample is a useful application. Contrary to other forms of biometric person identification used in forensic labs, automatic writer identification often allows for determining identity in conjunction with the intentional aspects of a crime, such as in the case of threat letters. This is a fundamental difference from other biometric methods, where the relation between the evidence material and the details of an offense can be quite remote [3].

In addition to the forensic application of writer identification and verification, [10] ink type recognition [11], security [12], forgery detection [13], and writer identification on medieval and historical documents [14] are also researched. Writer identification and verification of others languages include Chinese [15], Dutch [16], Arabic [4],[17-21], Farsi [22], and Greek [23].

Researchers used different types of features for writer identification. He et al. used contourlets [15], others used Gabor filters [24], Moment-Based features [25], graphemes [26], a sliding window to extract both local and global features [27], Directional Element Features (DEFs) [16], connected components contours, and a collection of local features [13].

To the best of the researchers' knowledge, only few researchers [4],[17-21] have addressed writer identification and verification of Arabic text. Bulacu et al. [4] used the IFN/ENIT dataset [28] which is limited to Arabic town and city names. Srihari and Ball [21] used a dataset of 10 different writers, each contributing 10 different full page documents in handwritten Arabic for a total of 100 documents. Using macro- and micro-features along with likelihood ratio computation, they reported 86% accuracy. Al-Ma'adeed et al. [19],[20] used edge-based statistical features to recognize Arabic handwritten words. They asked 100 volunteers to write 16 Arabic phrases and words 20 times and performed writer identification on collected samples. Some of the phrases scored a Top-10 result of > 90% accuracy, whereas shorter words scored around 50% accuracy. Gazzah and Ben Amara [17] used 2d-Discrete Wavelet Transforms (DWT) on a database of 180 letters collected from 60 writers where each writer wrote the letter three times. Their reported accuracy is around 95%. Finally, Al-Dmour and Zitar [18] presented a technique for feature extraction based on hybrid spectral-statistical measures (SSMs) of texture. Experiments were performed using Arabic handwriting samples from 20 different people and 90% correct identification was achieved.

In this work Arabic handwritten digits identification is addressed as a first step in a comprehensive research in the effort of writer identification and verification of Arabic handwritten text. The authors are not aware of any previous work in writer identification of Arabic handwritten digits. To our knowledge, only one work was performed on writer identification in Latin handwritten digits [13].

3. Features

In this work multiple types of features are used. Gradient, curvature, density, horizontal and vertical run length, stroke, and concavity shape features are implemented. A concise summary of these features is given below. Some of these features are classically known [3],[29], and has been successfully implemented by the authors in previous digit recognition tasks [30],[31]. Gradient features have been tweaked to better suit Arabic digits as explained next.

3.1. Gradient features

The digit image is divided into *n x m* grids with equal number of black pixels for each of *n* rows, and for each of *m* columns. The features of the individual grid segments are extracted.

The gradient features are computed by convolving two *3 x 3* Sobel operators with the binary image. These operators approximate the *x* and *y* derivatives in the image at a pixel position. The gradient of a centre pixel is computed as a function of its eight nearest neighbours. The vector addition of the operators' output is used to compute the gradient of the image. The direction of the gradient vector is used in the computation of the feature vector. The direction of the gradient can range from 0 to 2π radians. A sliding window of half a quadrant is used to estimate the histogram of gradient directions of the pixels in the window. Each histogram value corresponds to the count of each gradient direction in the sliding window. The sliding window overlaps with the previous window by 1/3 of the window range (i.e. 15 degrees). Starting at angle 0, the first half quadrant window extends from 0 to 45 degrees; the second quadrant extends from 30 to 75 degrees and so on. The overlapped sliding windows produce *12 x n x m* features representing the gradient feature vector (where *n* and *m* are the number of horizontal and vertical segments respectively). Fig. 1 shows an illustration of the Cartesian space with the first and second half quadrant windows highlighted.

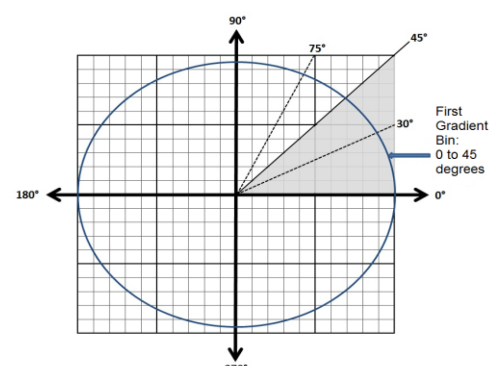

Fig. 1. First and second gradient feature bins.

3.2. Curvature features

The contour of the digit is extracted and encoded using Freeman chain codes shown in Fig. 2. The external angles between every two consecutive direction codes (i.e. two consecutive edges on the contour) are used to obtain the concave, convex and straight segments' features at different angles. Fig. 3 shows all the combinations of the used features. The left column is the feature label and the right columns are the directions assigned to the label. Labels 1 to 8 are subcategories of the concave features, Labels 9 to 16 are the subcategories of the convex features, and labels 17 to 20 are the straight line features. Due to the finer resolution used in the curvature features; the digit enclosing rectangle is divided into four quadrants. The number of features, of each curvature feature type, in each quadrant is estimated. A total of 80 features are extracted for the digit.

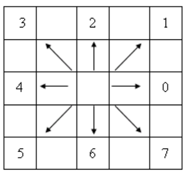

Fig. 2. Freeman chain codes relative to the center point.

3.3. Density features

The average density of the black pixels in each image segment is calculated and used as a feature. This feature contributes *n x m* features as the digit has *n x m* segments.

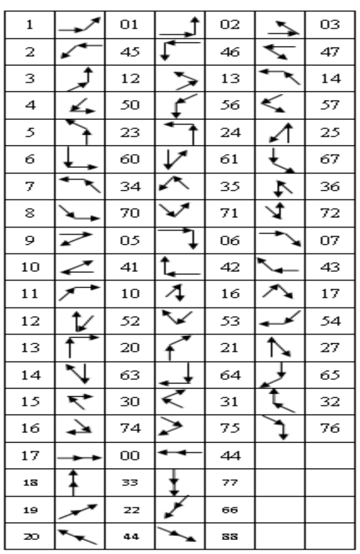

Fig. 3. Directions of concave, convex, and straight features.

3.4. Horizontal and vertical run lengths

The horizontal and vertical run lengths in each image segment are accumulated by adding the count of black horizontal and vertical lines that constitute a run of more than 2 pixels. This feature contributes *2 x n x m* features.

3.5. Stroke features

These features estimate the number of horizontal, vertical, left- and right-diagonal strokes in the image segments. Run lengths of horizontal, vertical, left- and right-diagonal black pixels across the image are first computed. From this information, the presence of strokes is determined by storing the maximum horizontal, vertical, left- and right-diagonal run length in each region. This feature contributes *4 x n x m* features.

3.6. Concavity shape features

These features are computed by convolving the image with a star like operator. This operator shoots rays in eight directions and determines what each ray hits. A ray can hit an image pixel or the edge of the image. Upward/downward, left/right pointing concavities are detected along with holes. The rules are relaxed to allow nearly enclosed holes (broken holes) to be detected as holes. These features contribute *5 x n x m* features.

4. Experimental results

Abdleazeem et al. described their Arabic Digits dataBase (ADBase) in [32]. ADBase is composed of 70,000 digits written by 700 participants. Each participant wrote each digit (from '0' to '9') ten times. Images size in pixels varies from 3 by 5 pixels for the smallest image and up to 140 by 29 pixels for the largest image. Fig. 4 shows samples of the ADBase. The database is partitioned into two sets for the purpose of digits recognition: a training set (60,000 digits to 6,000 images per class) and a test set (10,000 digits to 1,000 images per class). Writers of training and test sets are disjoint.

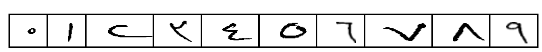

Fig. 4. Samples of ADBase.

In order to apply the database into writer identification, we divided the database into two sets: training set and testing set. The training set contains 49,000 digits (70% of the dataset), whereas the testing set contains 21,000 digits (30% of the dataset). For each writer, 70 random digits are selected for the training set (7 samples per digit for each writer), and 30 random digits are selected for the testing set (3 samples per digit for each writer). The training set is further divided into initial-training set and verification set for selecting the optimal number of grids for our features. Initial-training set contains 35,000 digits (71.4% of the training set, 5 samples per digit per writer). The verification set contains 14,000 digits (28.6% of the training set, 2 samples per digit per writer).

Nearest Neighbor (NN) and Nearest Mean (NM) are simple classifiers that are used to measure the effectiveness of the extracted features and the identifiability of Arabic handwritten digits. The nearest neighbor is computed using an Euclidean distance classifier. This model is considered as the writer class that matches most closely the obtained features vector of the unknown writer in Euclidean space. Writer identification researchers have preferred the use of distance and dissimilarity measures over statistical classifiers like Hidden Markov Models (HMM) and Support Vector Machines (SVM) mainly because of the nature of the writer identification problem [4],[33],[34]. The problem of writer identification usually involves large number of classes (i.e. writers) and few samples per class (i.e. digits per writer) compared to relatively few classes (i.e. number of distinct digits) and large number of samples per class (i.e. samples of images per digit) common in digit recognition scenarios.

The models for the NM classifier are taken as the mean of all the features of the training samples for each digit of each writer. This is done by averaging the features of 7 samples of each digit of each writer and using them as the feature models for the writers.

After smoothing the images, the above features are extracted for each set (e.g. for a *2 x 2* division, the concatenation of all of the features resulted in a 172-dimensional feature vector, viz. 48 gradients, 80 curvature, 4 density, 4 horizontal and vertical run lengths, 16 stroke, and 20 concavity features).

In order to estimate the optimal number of grid segments of the digit image, several experiments are conducted using divisions of *2 x 2* up to *8 x 8* on the initial-training and verification sets. The experimental results have shown that *5 x 5* divisions resulted in the highest recognition rate. Fig. 5 shows a sample of digit 9 (٩) divided into *5 x 5* divisions.

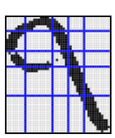

Fig. 5. Digit 9 (٩) divided into *5 x 5* divisions.

With *5 x 5* grid divisions, training and testing is performed on the ADBase. Using the NN classifier and the above features we tested Arabic handwritten digits identification and recognition using 21000 samples. Table 1 shows the top-1, top-5 and top-10 writer-identification performance results as well as digit classification results.

Table 1: Writer identification and digit recognition accuracy for each digit using NN

Digit	Accuracy of Writer Identification			Digit Recognition		
	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10
0 (٠)	2.86%	8.81%	14.19%	98.90%	99.86%	99.90%
1 (١)	4.29%	11.76%	16.48%	99.10%	99.71%	99.86%
2 (٢)	12.10%	28.05%	36.95%	99.43%	99.90%	99.90%
3 (٣)	19.81%	38.19%	47.67%	98.90%	99.48%	99.48%
4 (٤)	17.10%	33.81%	42.76%	99.10%	99.71%	99.86%
5 (٥)	14.14%	29.52%	38.81%	99.43%	99.81%	99.81%
6 (٦)	8.86%	23.19%	30.76%	99.29%	99.86%	99.86%
7 (٧)	13.10%	32.90%	43.90%	99.76%	99.90%	99.95%
8 (٨)	15.52%	33.95%	43.52%	99.71%	99.81%	99.95%
9 (٩)	15.29%	31.81%	41.00%	98.76%	99.62%	99.81%
Total	12.30%	27.20%	35.60%	99.24%	99.77%	99.84%

For the NM, the feature vectors for the training set for each digit and each writer are averaged. This reduces the number of feature vectors for training set from 49,000 training vectors into 7,000 averaged vectors (700 writers, 10 average digits per writer). Table 2 shows the results for each digit. The table shows that even though averaging the training features has increased the top-1 accuracy results for writer identification, it has failed to do so in the top-5 and top-10 categories. This is somewhat expected since there is only one training vector for each digit per writer instead of seven vectors, and hence reducing the possibility of getting a hit in the top-5 and top-10 lists. In addition, averaging vectors have reduced digit recognition rate as expected since many inter-digit variation decreases (e.g. an Arabic three digit would look more like an Arabic two digit).

Table 2: Writer identification and digit recognition accuracy for each digit using NM

Digit	Accuracy of Writer Identification			Digit Recognition (%)		
	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10
0 (٠)	4.05%	11.14%	17.33%	98.86%	99.38%	99.67%
1 (١)	4.33%	11.29%	16.00%	96.81%	98.33%	98.62%
2 (٢)	14.67%	31.19%	40.86%	98.10%	99.33%	99.52%
3 (٣)	22.05%	42.10%	50.95%	98.24%	99.19%	99.33%
4 (٤)	17.52%	33.48%	41.95%	98.86%	99.48%	99.62%
5 (٥)	17.90%	35.71%	45.29%	99.29%	99.48%	99.48%
6 (٦)	11.57%	24.71%	33.86%	98.86%	99.62%	99.76%
7 (٧)	17.71%	37.62%	48.24%	99.57%	99.81%	99.86%
8 (٨)	18.81%	38.62%	49.24%	99.24%	99.52%	99.57%
9 (٩)	17.71%	34.48%	44.76%	98.67%	99.33%	99.33%
Total	14.63%	30.03%	38.85%	98.65%	99.35%	99.48%

In several experiments the combination of similar digits are tested and they did not improve the identification or recognition rates. In order to combine the discriminatory power of each digits, the extracted features for each group of digits (0 to 9) is concatenated to form one feature vector. This is implemented simply by the concatenation of the features of the different digits as the database consists only of isolated digits. These concatenated feature vectors are used in the analysis using one classifier. This produced 2100 feature vectors for testing (700 writers, 3 feature vectors per writer) along with 4900 concatenated training vectors (700 writers, 7 feature vectors per writer) for the k-NN classifier and 700 averaged and concatenated training vectors (700 writers, 1 feature vector per writer) for NM classifier. Since each digit' feature vector is compared to its corresponding digit feature vector in the training set, we consider this approach to be text-dependent writer identification. Table 3 shows a summary of the writer identification results for the text-dependent approach. Results for each feature groups, i.e. gradient, curvature, and concavity, is shown as well as results for all features combined. Density features, horizontal and vertical run lengths, stroke features, and concavity shape features are all concatenated together and called 'concavity' features due to their relatively small size.

Table 3: Text-Dependent Writer Identification

	NN Classifier			NM Classifier		
	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10
Gradient	66.19%	79.76%	84.14%	80.33%	89.76%	92.38%
Curvature	60.19%	78.71%	84.86%	74.90%	89.14%	92.62%
Concavity	65.71%	79.86%	83.76%	82.67%	91.48%	94.19%
All Features	69.52%	81.67%	85.81%	81.33%	90.67%	92.86%

Finally, we compare each digit in the testing set against all digits in the training set for each writer and store its writer identification rank. We do this for all the digits (0-9) for that specific writer, and then we add the rank for each writer and select the most probable writer, and hence we consider this approach to be text-independent writer identification. Table 4 shows a summary of the writer identification results for the text-independent approach.

Table 4: Text-Independent Writer Identification

	NN Classifier			NM Classifier		
	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10
Gradient	84.62%	93.62%	95.90%	85.24%	93.24%	95.14%
Curvature	67.38%	84.29%	89.67%	67.10%	83.76%	89.24%
Concavity	86.43%	94.67%	96.19%	86.71%	93.90%	95.81%
All Features	88.14%	94.81%	96.48%	87.76%	94.10%	96.14%,

5. Conclusion

The presented work addresses the identifiability of Arabic handwritten digits. Nearest Mean and k-Nearest Neighbors are used for classification. In addition to digit identifiability, the paper presents digit recognition. Gradient, curvature, density, horizontal and vertical run lengths, stroke, and concavity features are used. A database of Arabic handwritten digits written by 700 different writers is used in the analysis.

A numbers of experiments are carried out to select the optimal number of digit divisions for the feature extraction phase. Combining all digits and finding the NN provided the best accuracy in text-independent writer identification with top-1 result of 88.14%, top-5 result of 94.81%, and top-10 results of 96.48%. The analysis of the results indicates that Arabic digits 3 (٣), 4 (٤), 8 (٨), and 9 (٩) are more identifiable than other digits while Arabic digit 0 (٠) and 1 (١) are the least identifiable. K-NN provided best accuracy for digit recognition with top-1 result of 99.24%, top-5 result of 99.77%, and top-10 results of 99.84%, with only 34 erroneously classified digits out of 21,000 test digits in the top-10 results.

These encouraging results demonstrate the discriminability of Arabic digits for writer identification. The researchers are extending these features for writer identification using Arabic handwritten text.

Acknowledgment

The authors would like to thank King Fahd University of Petroleum and Minerals for supporting this research and providing the computing facilities.

References

[1] R.R. Bradford and R.B. Bradford, *Introduction to Handwriting Examination and Identification*, Burnham, Inc., 1992.

[2] S.N. Srihari, S. Cha, H. Arora, and S. Lee, "Individuality of Handwriting," *Journal of Forensic Sciences (JOFS)*, vol. 47, Jul. 2002.

[3] L. Schomaker and M. Bulacu, "Automatic writer identification using connected-component contours and edge-based features of uppercase Western script," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 26, 2004, pp. 787-798.

[4] M. Bulacu, L. Schomaker, and A. Brink, "Text-Independent Writer Identification and Verification on Offline Arabic Handwriting," *Ninth International Conference on Document Analysis and Recognition, 2007.* pp. 769-773.

[5] L. Schomaker, M. Bulacu, and K. Franke, "Automatic writer identification using fragmented connected-component contours," *Ninth International Workshop on Frontiers in Handwriting Recognition, 2004.* pp. 185-190.
[6] K. Franke and M. Koppen, "A computer-based system to support forensic studies on handwritten documents," *International Journal on Document Analysis and Recognition*, vol. 3, May. 2001, pp. 218-231.
[7] S. Srihari and G.R. Ball, "Comparison of statistical models for writer verification," *Proceedings of SPIE*, San Jose, CA, USA: 2009, pp. 72470E-72470E-8.
[8] A. Brink, M. Bulacu, and L. Schomaker, "How much handwritten text is needed for text-independent writer verification and identification," *19th International Conference on Pattern Recognition, 2008.* pp. 1-4.
[9] A. Schlapbach and H. Bunke, "Using HMM Based Recognizers for Writer Identification and Verification," *Ninth International Workshop on Frontiers in Handwriting Recognition*, Tokyo, Japan: , pp. 167-172.
[10] R. Niels, L. Vuurpijl, and L. Schomaker, "Automatic allograph matching in forensic writer identification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 21, 2007, pp. 61-81.
[11] K. Franke, O. Bünnemeyer, and T. Sy, "Ink Texture Analysis for Writer Identification," *International Workshop on Frontiers in Handwriting Recognition,* Los Alamitos, CA, USA, 2002, p. 268.
[12] Ming-Yen Tsai and Leu-Shing Lan, "Online Writer Identification Using The Point Distribution Model," *IEEE International Conference on Systems, Man and Cybernetics*, 2005, pp. 1264-1268.
[13] G. Leedham and S. Chachra, "Writer identification using innovative binarised features of handwritten numerals," *Seventh International Conference on Document Analysis and Recognition, 2003.* pp. 413-416 vol.1.
[14] M. Bulacu and L. Schomaker, "Automatic Handwriting Identification on Medieval Documents," *14th International Conference onImage Analysis and Processing*, 2007, pp. 279-284.
[15] Z. He, Y.Y. Tang, B. Fang, J. Du, and X. You, "A Novel Method for Off-line Handwriting-based Writer Identification," *International Conference on Document Analysis and Recognition,* Los Alamitos, CA, USA, 2005, pp. 242-246.
[16] M. Bulacu and L. Schomaker, "A comparison of clustering methods for writer identification and verification," *Proceedings. Eighth International Conference on Document Analysis and Recognition, 2005*, pp. 1275-1279 Vol. 2.
[17] S. Gazzah and N. Ben Amara, "Arabic Handwriting Texture Analysis for Writer Identification Using the DWT-Lifting Scheme," *Ninth International Conference on Document Analysis and Recognition, 2007*, pp. 1133-1137.
[18] A. Al-Dmour and R.A. Zitar, "Arabic writer identification based on hybrid spectral-statistical measures," *J. Exp. Theor. Artif. Intell.*, vol. 19, 2007, pp. 307-332.
[19] S. Al-Ma'adeed, A. Al-Kurbi, A. Al-Muslih, R. Al-Qahtani, and H. Al Kubisi, "Writer identification of Arabic handwriting documents using grapheme features," *IEEE/ACS International Conference on Computer Systems and Applications, 2008*, pp. 923-924.
[20] S. Al-Ma'adeed, E. Mohammed, and D. Al Kassis, "Writer identification using edge-based directional probability distribution features for arabic words," I*EEE/ACS International Conference on Computer Systems and Applications,* 2008, pp. 582-590.
[21] S. Srihari and G. Ball, "Writer Verification of Arabic Handwriting," *The Eighth IAPR International Workshop on Document Analysis Systems, 2008.* pp. 28-34.
[22] F. Shahabi and M. Rahmati, "Comparison of Gabor-Based Features for Writer Identification of Farsi/Arabic Handwriting," *Proceedings of 10th IWFHR*, 2006, pp. 550, 545.
[23] E. N and V, "Morphological waveform coding for writer identification," *Pattern Recognition*, vol. 33, Mar. 2000, pp. 385-398.
[24] C. Liu, R. Dai, and Y. Liu, "Extracting individual features from moments for Chinese writer identification," *International Conference on Document Analysis and Recognition,* Los Alamitos, CA, USA, 1995, p. 438.
[25] Soe Hae Mar and Ni Lar Thein, "Myanmar Character Identification of Handwriting Between Exhibit and Specimen," *Proceedings. 6th Asia-Pacific Symposium on Information and Telecommunication Technologies, 2005*, pp. 95-98.
[26] A. Bensefia, T. Paquet, and L. Heutte, "Information Retrieval Based Writer Identification," *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, 2003, p. 946.
[27] H.B. Andreas Schlapbach, "Using HMM Based Recognizers for Writer Identification and Verification," Oct. 2004.
[28] M. Pechwitz, S. Snoussi Maddouri, V. M?rgner, N. Ellouze, and H. Amiri, "IFN/ENIT - Database of Handwritten Arabic Words," *7th Colloque International Francophone sur l'Ecrit et le Document,* Hammamet, Tunis: 2002.
[29] J. Favata and G. Srikantan, "A multiple feature/resolution approach to handprinted digit and character recognition," *International Journal of Imaging Systems and Technology*, vol. 7, 1996, pp. 311, 304.
[30] S.M. Awaidah and S.A. Mahmoud, "A multiple feature/resolution scheme to Arabic (Indian) numerals recognition using hidden Markov models," *Signal Process.*, vol. 89, 2009, pp. 1176-1184.
[31] S. Mahmoud and S. Awaida, "Recognition of Off-Line Handwritten Arabic (Indian) Numerals Using Multi-Scale Features and Support Vector Machines vs. Hidden Markov Models," *Arabian Journal for Science and Engineering (AJSE)*, vol. 34, Oct. 2009, pp. 429-444.
[32] S. Abdleazeem and E. El-Sherif, "Arabic handwritten digit recognition," *International Journal on Document Analysis and Recognition*, vol. 11, Dec. 2008, pp. 127-141.
[33] X. Wang, X. Ding, and H. Liu, "Writer Identification Using Directional Element Features and Linear Transform," *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*, 2003, p. 942.
[34] B. Zhang, "Handwriting pattern matching and retrieval with binary features," State University of New York at Buffalo, 2003.