Recognition of Arabic Handwritten Amount in Cheque through Windowing Approach

Mowaffak O.A. Albaraq
Al-maafeer Community College and
Computer Science and Engineering
National University,
Republic of Yemen, Taiz
Contact No.(00967-773922007)

S.C.Mehrotra
Department of Computer Science and Information Technology,
Dr. BAMU University,
Aurangabad – 431001,
Maharshtra. State, India.

ABSTRACT
Arabic language is a semantic language that has differences when compared to English language. We are dealing with the handwritten Arabic Amount from cheques of Arabic banks . In this paper we proposed a windowing technique for the segmentation of the numerical amount, followed by an efficient moment invariants for features extraction . A maximum and minimum points technique used to isolate the Arabic (Hindi Digits) numerals. The feature vectors are grouped for each digit and Artificial Neural Network (ANN), is applied for the classification and recognition. This approach resulted in providing 99.5% of recognition rate.

Keywords
Windowing approach, Moment Invariants, Features Extraction, Handwritten Arabic Checks , ANN and OCR.

1. INTRODUCTION
Arabic handwritten Amount (Hindi digits) recognition is useful in a large variety of banking, business applications, postal zip code reading, and data entry applications. Arabic words are cursive both in printed and handwritten forms, Arabic (Hindi digits) numerals are not cursive[1],[2],[3]. This simplifies the processing, as there is no need to segment the digits as in the case with Arabic text. Indian numerals are used in Arabic writing while Arabic numerals are used in Latin languages the Arabic numerals 0 to 9 and its corresponding Indian numerals are shown in figure 1 [4],[5]. We addressing the some issues for the Arabic (Hindi digits) numerals as follows:

1. Arabic numerals can be written form any orders middle wright to left or left to wright .
2. Arabic literals amounts has more than literal amount in Latin or French as we got the 52 general classes have used in the 22 Arabic countries .
3. Most the Arabic checks has a common fields which is specific the legal amount field courtesy amount filed, the symbol of the banks the name of the country field and the signatures two or four field.

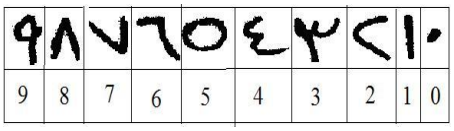

Figure 1 : Arabic Numerals with its corresponding Indian Numerals.

4. The similarity between eight and seven digit and two with six digits make the ambiguous of the recognition or make it more difficult forth more (see figure 1.),[3],[7].
5. Arabic handwritten legal amount contents a coma after each three digits and closed by hash sign (see figure 3).

The remaining of this paper will discuss the system architecture in section 2 and the phases for the recognition are presented in section 3, section 4 deals with the results and experiment, conclusion is presented in section 5.

2. SYSTEM ARCHITECTURE
Any OCR system started with data acquisition. The architecture of numeral Arabic (Hindi digits) recognition system is presented in figure 2.

3. PHASES FOR RECOGNITION
3.1 Data Acquisition
The data are collected from cheques of Arabic bank's written by 100 subjects, all cheques belong to central bank of Yemen[5].

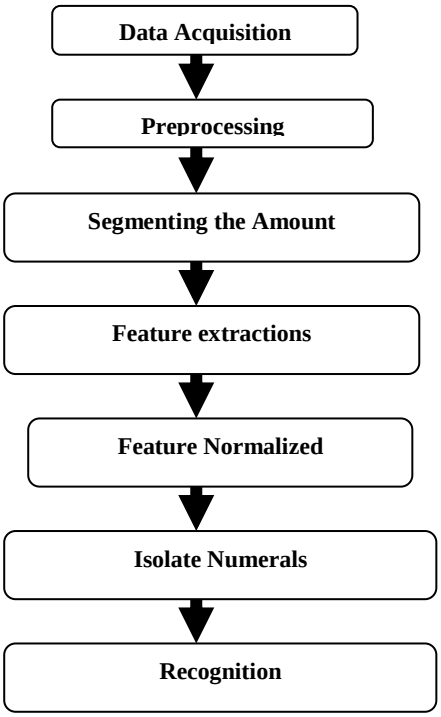

Figure 2. System Architecture

Any cheque contains many fields like date field ,number of the cheque, the symbols, the account number, the owner account, the text amount in words, numerical amount and three fields for signatures . our work focuses only on extracted the numerical amount from such cheque for the recognition purpose [6],[7].

3.2 Preprocessing

Second step deals with the cheque and extract the field of numerals amounts. The following steps are used to get the pure data:-

3.1.1 Noise detection and removal

Noises must be removed from such legal amounts and remove the dark spots, ink, coma and extra line behind the Amount.

3.1.2 Slop correction

Handwritten digits must be normalized to the horizontal position.

3.2 Segmentation Legal Amount

We applied windowing method for segment the amount into atomic digits along with storing each as file images. The scanning is done from right to left. Each numeral has unique specifications which is extracted and stored with the appropriate width's and long's.

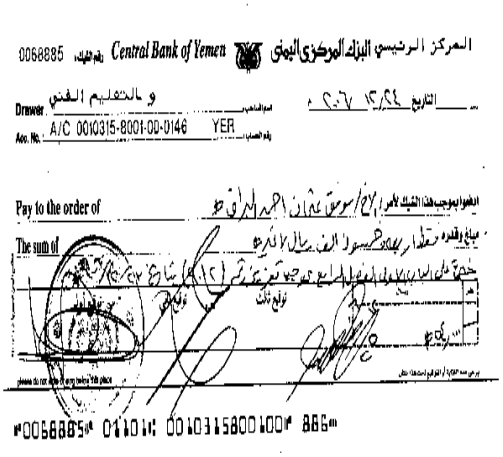

Figure 3 : Arabic Handwritten cheque Of the central bank of Yemen.

Figure 4. Amount Extracted from the check

3.3 Feature Extraction

The Feature extraction are done through seven invariant moments as given by Hu[5]. We applied these formulas to extract features from all segments stored as files images. Important prosperities of these moments that they are invariant under reflection, rotation and scaling. The regular moments are defined as:

$$M_{pq} = \iint X^p Y^q f(X,Y) \qquad (1)$$

Where p, q = 0,1,2,….and M*pq* is the (p+q)th order moment of the continuous image function f(x,y). If the image is represented by a discrete function, integrals are replaced by summations. Equation (1) can be written as follows:

$$M_{pq} = \sum_x \sum_y X^p Y^q f(X,Y) \qquad (2)$$

The central moments of μpq are evaluated with the expression as given blue

$$\mu_{pq} = \sum_X \sum_Y (X - \bar{X})^p (Y - \bar{Y})^q f(X,Y) \qquad (3)$$

Where $\bar{X} = m_{10} / m_{00}$, $\bar{Y} = m_{01} / m_{00}$ which are the centroid of the image. The central moments of order up to 3 are as follows :

$$\begin{aligned}
\mu_{00} &= m_{00} \\
\mu_{10} &= 0 \\
\mu_{01} &= 0 \\
\mu_{11} &= m_{11} - \bar{Y} m_{10} \\
\mu_{20} &= m_{20} - \bar{X} m_{10} \qquad (4) \\
\mu_{02} &= m_{02} - \bar{Y} m_{01} \\
\mu_{30} &= m_{30} - 3\bar{X} m_{20} + 2\bar{X}^2 m_{10} \\
\mu_{03} &= m_{03} - 3\bar{Y} m_{02} + 2\bar{Y}^2 m_{01} \\
\mu_{21} &= m_{21} - 2\bar{X} m_{11} - \bar{Y} m_{20} + 2\bar{X}^2 m_{01} \\
\mu_{12} &= m_{12} - 2\bar{Y} m_{11} - \bar{X} m_{02} + 2\bar{Y}^2 m_{10}
\end{aligned}$$

The normalized central moment of order (p+q) is defined as ηpq=μpq/ μ00γ for p,q=0,1,2

Where $\gamma = \frac{(p+q)}{2} + 1$ (2)

For (p + q) = 2,3

Based on normalized central moments, A set of seven moment invariants can be derived as follows:

In this work feature has been extracted from each window, (width 20 pixels, height 80 pixels). The expiration given by equation (5) which used to evaluate 7 center moment invariants i.e(Φ1- Φ7) which are used as features. In experiment , the moment invariants are evaluated by log of the absolute value of the moment of each of the image to be in numeral values.

3.4 Feature Representation

The figure (6-a) to (6-g) show the plots of (Φ1- Φ7) as a function of windows movement for the sentence 1 given in figure 4. The following observations can be made from the plots.

1. The polynomial minima represent the end of the word. This feature is used to identify the beginning and end of words in a sentence. The figure seven shows the values of Φ1 for each word after separating from the sentence.
2. After separation of the word, the feature related to the a word can be represented by polynomial of order n . The coefficient of these polynomials can be used as a feature for fore recognition of the digits . the figure 5 along with equation 5 shows the fit of the polynomial of order 7 along with values of coefficient [5].

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^{\ 2}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \quad (5)$$
$$(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] +$$
$$4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 +$$
$$(3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Figure 5: Seventh centers moment invariants i.e(Φ1- Φ7)

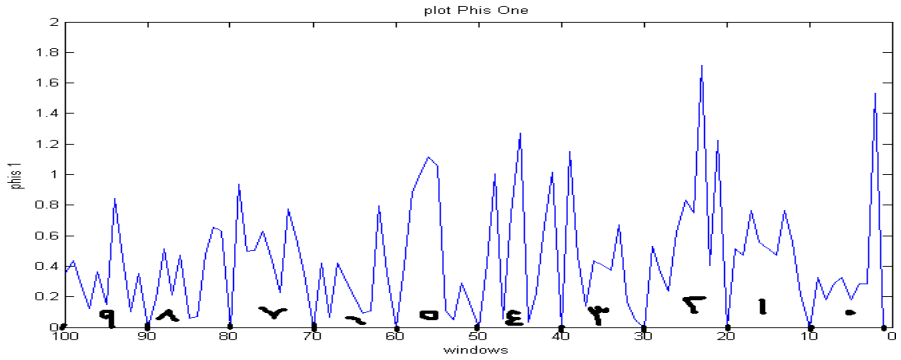

Figure 6-a

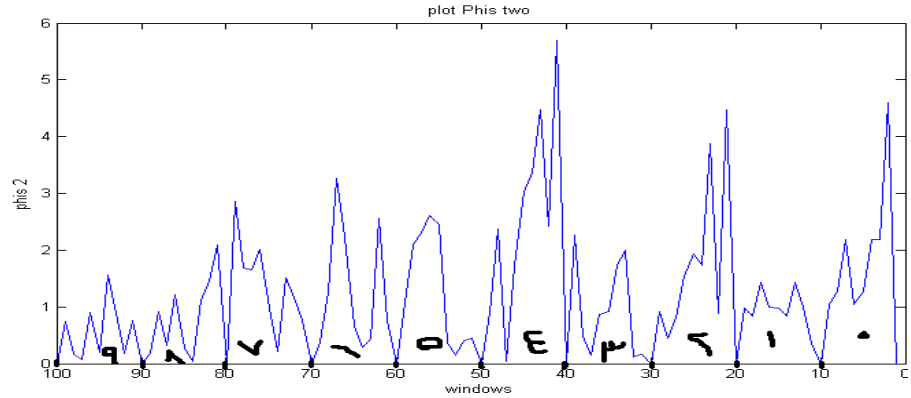

Figure 6-b

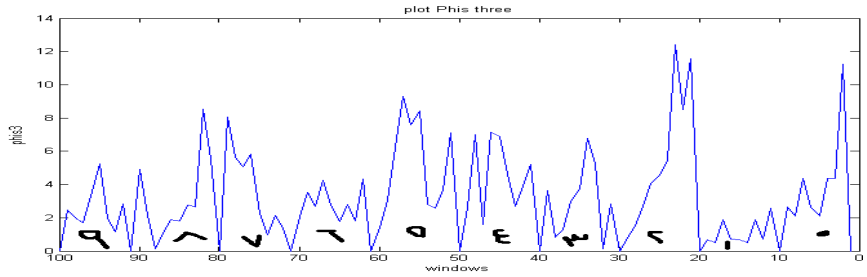

Figure 6-c

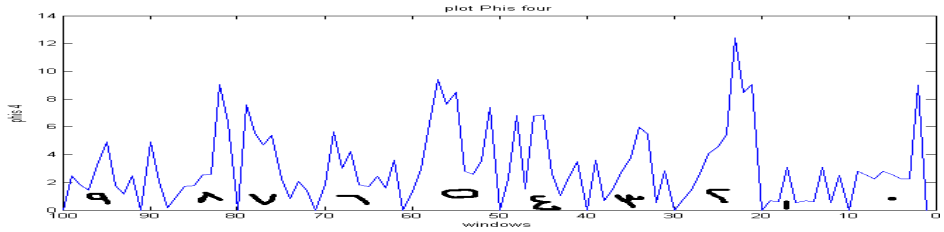

Figure 6-d

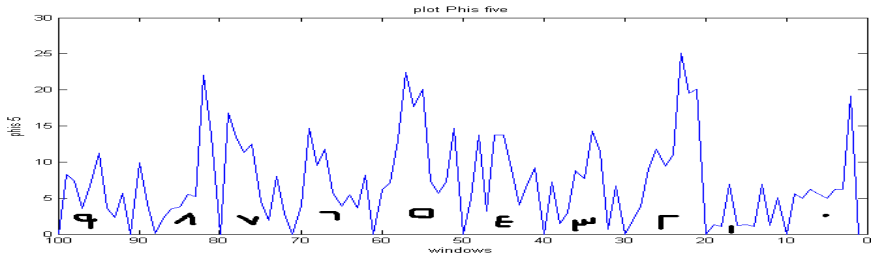

Figure 6-e

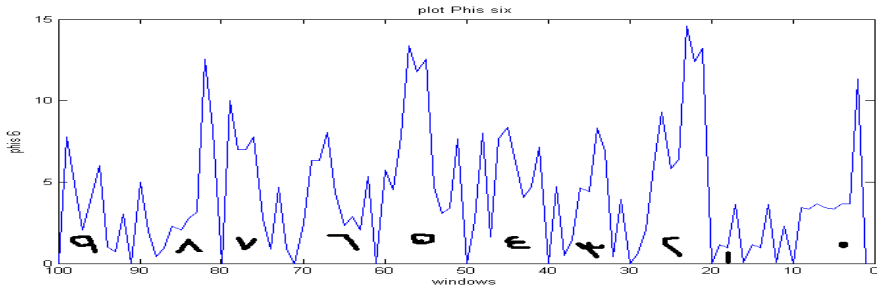

Figure 6-f

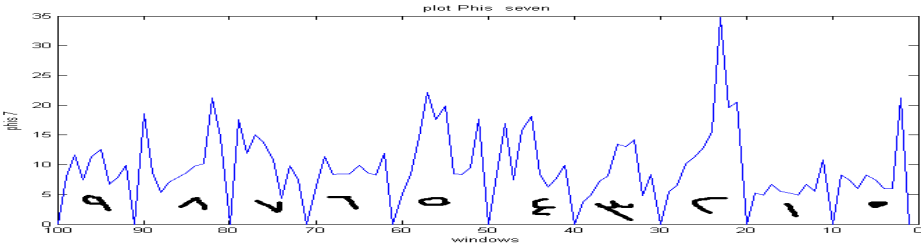

Figure 6-g

3.5 Recognition

For recognition the Arabic amount (Hindi digits), we applied Artificial Neural Network technique (ANN), to recognize the Arabic numerals (Hindi digits).The setup of neural network deals with one input layer, one hidden layer and one output layer. The features of the numeral are given to input layer, after passing through the hidden layer the recognized digit is at the output layer. The result is shown in table II, table III and table IV respectively. The difference between epoch values effecting those results which obtained with time elapsed[15]. 70% of data are used for training and remaining specified for testing purpose. The recognition rate is increased when we increased the epoch setting into 90000 and with the elapsed time 967.55 second. The Features classified using above classifier and the result achieved is 99.5%.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Experiments have been performed to test the above system. The developed Arabic Handwritten Legal Amounts (Hindi digits), numerals has been tested using randomly selected Amounts had written on Arabic cheques.

Table II. Epoch 60000Elapsed time 612.25 sec

Nu5	*β1*	*β 2*	*β 3*	*B4*	*β 5*	*β 6*	*B7*	*B8*	*B9*	*B0*
α1	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α 2	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α3	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α 4	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
α 5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
α 6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
α 7	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00
α 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00
α 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
α 0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

The system is designed in Matlab 7.0 for the recognition of Arabic Handwritten Legal Amounts. The system developed is a single font, one size system. The system gained 99.5% successful recognition rate with elapsed time 16 minutes for more information evidence in tables II,III,IV and V respectively. Such that in table II when Epoch 60000 and Elapsed time is 612.25 the recognition rate was 99.1% seconds ,when the epoch 85000 elapsed time is 975.55 second the recognition rate improved and we reached the highest recognition rate when the epoch was 90000 and elapsed time is 1050.50 second[14] .

Table III. Epoch 85000 Elapsed time 976.55 se.

Nu5	*β1*	*β 2*	*β 3*	*B4*	*β 5*	*β 6*	*B7*	*B8*	*B9*	*B0*
α1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α 2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α 4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
α 5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
α 6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
α 7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
α 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
α 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
α 0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table IV. Epoch 900000 Elapsed time 1050.56

Nu5	*β1*	*β 2*	*β 3*	*B4*	*β 5*	*β 6*	*B7*	*B8*	*B9*	*B0*
α1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α 2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α 4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
α 5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
α 6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
α 7	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
α 8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
α 9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
α 0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

5. CONCLUSION

In this paper, an attempt is made to apply a different techniques based on a sliding window for segmentation the Amounts into atomic digits, moment invariants are used for features extraction. Plotting and smoothing for representation the desired amount, separation the amount into it's digits using minima and maxima points techniques. And finally for recognition we applied ANN technique to classified statistical features obtained from moment invariants (ϕ_1-ϕ_7). The system was achieved 99.5% successful rate for the ANN technique. We conclude that our system enhancement the recognition rate. Advantages of this method can be categorized as follows[13]:

- *High power of the extracted features*
- *Easy implementation*
- *Very low feature extraction time*
- *Very low computational overhead*

In future work, we will develop an automatic Arabic Recognition System for recognition of different fonts and size along with enhancing rate.

6. REFERENCES

[1] Sabri A. Mahmoud "Arabic (Indian) Handwritten Digits Recognition Using Gabor-based Features" ©2008 IEEE.

[2] A. Harifi and A. Aghagolzadeh, "A New Pattern for Handwritten Persian/Arabic Digit Recognition," Proceedings of World Academy of Science, Engineering and Technology, vol. 3, Jan 2005, pp. 130-133.

[3] Alotaibi, Y. A., Spoken Arabic Digits Recognizer Using Recurrent Neural Networks, International Symposium on Signal Processing and Information Technology- ISSPIT, 195-199, Rome, Italy, 2004.

[4] .Al-Taani Ahmad"Recognition of On-line Handwritten Arabic Digits Using Structural Features and Transition Network ", Informatica 32 (2008) 275–281 275.

[5] Mowaffak Albaraq and S.C. Mehrotra, "Handwritten Arabic Text Recognition System using Window based Moment Invariant Method", International JARCS, v2 No. 1, Jan. –Feb, 2011, 599-608.

[6] L. S. Meslati and M. Sellami, "A Hybrid Approach For Arabic Literal Amounts Recognition ",The Arab. Jou. for Sci. and Eng. , Vol. 29, No. 2B

[7] Y. Al-Ohali , M. Cheriet andd C. Suena "Databases for

recognition of handwritten Arabic cheques" Jou. of the Patt. Rec. Society , (2003).

[8] R.C Gonzalez and R.E. Woods, " Digital Image Processing ", Prentice Hall, 2002 .

[9] M.B. Menhadj, " Neural Networks ", Amirkabir Univ. Publs

[10] Ahmad T. Al-Taani, "An Efficient Feature Extraction Algorithm for the Recognition of Handwritten Arabic Digits " , International Journal of Computational Intelligence, 2006

[11] J. Sadri C.Y. Suen and T. D Bui "Application of Support Vector Machine for Recognition Of Handwritten Arabic/Persian Digits" K.N., Toosi Univ. of Tech., Iran Feb. 2003. vol 1.

[12] Sabri A. Mahmoud, "Arabic (Indian) Handwritten Digits Recognition Using Gabor-based Feature" 978-1-4244-3397-1/08/$25.00 ©2008 IEEE.

[13] Mohsen Biglari, Faezeh Mirzaei, Jalil Ghavidel Neycharan, "Persian/Arabic Handwritten Digit Recognition Using Local Binary Pattern", International Journal of Digital Information (IJDIWC), 2014.

[14] Nibaran Das, Ayatullah Mollah, Sudip Saha, Syed S. Haque, "Handwritten Arabic Numeral Recognition using a Multi Layer Perceptron", National Conference on Recent Trends in Information Systems (2006) 200-203.

[15] Gita Sinha and Dr. Jitendra kumar "Arabic numeral Recognition Using SVM Classifier", International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-2, Issue-5).

[16] Sabri A. Mahmoud and Sameh M. Awaida, "Recognition Of Off-Line Handwritten Arabic (Indian) Numerals Using Multi-Scale Features And Support Vector Machines Vs. Hidden Markov Models", The Arabian Journal for Science and Engineering, Volume 34, Number 2B, October 2009.