Comparative Study of the Use of Geometrical Moments for Arabic Handwriting Recognition

Maâmar Kef  
Department of Computer Sciences  
University Hadj Lakhdar  
Batna, Algeria  
lm_kef@yahoo.fr

Leila Chergui  
Department of Computer Sciences  
University Larbi Ben Mhidi  
Oum el Bouaghi, Algeria  
pgleila@yahoo.fr

Salim Chikhi  
Department of Computer Sciences  
University Mentouri  
Constantine, Algeria  
schikhi@yahoo.com

Abstract—Moments and functions of moments have been employed as pattern features in numerous applications to recognize two-dimensional image patterns. These pattern features extract global properties of the image such as the shape area, the center of the mass, the moment of inertia, and so on. This paper shows the use of different moments to extract features from offline Arabic words. Invariants moment of Hu, Zernike moments, Pseudo Zernike moments, Tchebichef moments, and Legendre moments have been applied to the IFN/ENIT database with a neural network classifier and the results have been compared. Our results show that pseudo Zernike moments yields the best recognition accuracy of 89%.

Keywords- Arabic recognition, Invariant moments, Tchebichef moments, Multi Layer Perceptron.

I. INTRODUCTION

During the past decade, remarkable progress has been achieved in the field of handwritten word recognition. Many papers dealing with applications of handwritten word recognition to automatic reading of postal addresses, bank checks and forms (invoices, coupons and revenue documents, etc.) have been published. However, most of the published work deals with the recognition of Latin and Chinese scripts. The progress in Arabic script recognition has been slow mainly due to the special characteristics of these scripts.

Generally holistic or character based methods have been used for handwritten word recognition. In the former approach, a word is treated and identified as an entity. In the second approach, a word is considered as a sequence of smaller components-like characters or graphemes. The word is identified by extracting and recognizing its constituent components. The choice of a specific approach for a particular application is strongly influenced by the size of the available vocabulary (lexicon). The holistic approach can be used if the size of the vocabulary is small (as would be the case in a typical application such as the recognition of the legal amount in cheques provided the words in the phrase are well isolated). The character-based approach is generally the preferred method for recognition applications that involve large-size vocabularies. In applications such as recognition of city and street names from the address blocks, both approaches are feasible. The function moments have been used as shape descriptors in a variety of applications in image analysis, like visual pattern recognition, object classification, template matching, edge detection, pose estimation, robot vision, and data compression. In all these applications, geometric moments and their extensions in the form of radial and complex moments have played important roles in characterizing the image shape, and in extracting features that are invariant with respect to image plane transformations.

The contribution of the paper is two-fold. Firstly, the use of each class of moments with a neural classifier is presented. Secondly, a comparative study between the efficiency of each type of moment is established.

The remaining part of the paper is organized as follows. In section 1 we will present the Arabic script. Section 2 focuses on summarize the used moments. Section 3 describes the architecture of the Multi Layer Perceptron. Experiments and results are discussed in section 5 and conclusions are drawn in the last section of this paper.

II. ARABIC RECOGNITION

The recognition of Arabic script has many applications such as mail sorting, bank check reading, and, more recently, the recognition of historical manuscripts.

The Arabic script evolved from a type of Aramaic, with the earliest known document dating from 512 AD. The Aramaic language has fewer consonants than Arabic, so new letters were created around the 7th century by adding dots to existing letters [3]. Arabic is a language written by more than 200 million people, the most obvious characteristic of the Arabic language are:

- Arabic script is inherently cursive.
- It is written from right to left.
- The Arabic alphabet has 28 basic letters (figure 1).
- The letters can have four different shapes, depending on their position in the word (beginning, middle, end or alone).
- Some character combinations form new ligature shapes which are often font dependent.
- Many Arabic characters have dots which are positioned at a suitable distance above or below the letter body. Dots can be single, double, or triple.
- Spacing may separate not only words but also certain characters within a word forming sub-words.
- Some Arabic letters may have a zig-zag like stroke called Hamza.
- Arabic characters are connected on an imaginary line called baseline.
- Some characters contain closed loops.

<table>
<thead>
<tr>
<th>Beginning</th>
<th>Middle</th>
<th>End and attached</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Example of a figure caption. (figure caption)

Compared to Latin script where a lot of research work is done, the number of work for Arabic script recognition is quite limited, this is due to:

- Difficulties attached to its shape, we give as examples; ligatures (character formed by combining two or more letters in an accepted manner), overlaps, the several of fonts.
- The absence of the database for effecting tests and validations, the only available database is the IFN/ENIT database [8].

III. MOMENTS

We have used five class of geometrical and orthogonal moment:

A. Hu moments

Moment invariants used by Hu [11] have been utilized to build the feature space for MLP classifier. Using nonlinear combinations of geometric moments, Hu derived a set of invariant moments which have the desirable property of being invariant under image translation, scaling and rotation. The central moments, which are invariant under any translation, are defined as:

\[
M_{pq} = \int\int_{-\infty}^{\infty} (x-x_0)(y-y_0)^p f(x, y)dx dy
\]  

(1)

Where:

\[
\bar{x} = \frac{\overline{M_{10}}}{M_{00}}, \quad \bar{y} = \frac{\overline{M_{01}}}{M_{00}} \quad \text{and}
\]

\[
\overline{M}_{pq} = \int\int_{-\infty}^{\infty} x^p y^q f(x, y)dx dy
\]

(2)

However, for images, the continuous image intensity function \(f(x, y)\) is replaced by a matrix, where \(x\) and \(y\) are the discrete locations of the image pixels. The integrals in equations (3) and (4) are approximated by the summations:

\[
M_{pq} = \sum_{x=0}^{m} \sum_{y=0}^{n} (x-x_0)(y-y_0)^p f(x, y)dx dy
\]  

(3)

\[
\overline{M}_{pq} = \sum_{x=0}^{m} \sum_{y=0}^{n} x^p y^q f(x, y)dx dy
\]  

(4)

Where \(m\) and \(n\) are the dimensions of the image. The set of moment invariants that has been used by Hu are given by:

\[
\phi_1 = M_{20} + M_{02}
\]  

(5)

\[
\phi_2 = (M_{20} - M_{02}) + 4M_{11}^2
\]  

(6)

\[
\phi_3 = (M_{30} - 3M_{12})^2 + (3M_{21} - M_{03})^2
\]  

(7)

\[
\phi_4 = (M_{30} + M_{12})^2 + (M_{21} + M_{03})^2
\]  

(8)

\[
\phi_5 = (M_{30} - 3M_{12})(M_{30} + M_{12})
\]

\[
[((M_{30} + M_{12})^2 - 3(M_{21} + M_{03})^2)] + (3M_{12} - M_{03})
\]

(9)

\[
\phi_6 = (M_{20} - M_{02})[(M_{30} + M_{12})^2 - (M_{21} + M_{03})^2]
\]

\[
4M_{11}(M_{30} + M_{12})(M_{21} + M_{03})
\]

(10)

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

\[
3(M_{21} + M_{03})^2] + 3(M_{21} - M_{03})(M_{21} + M_{03})^2
\]

(11)

These functions can be normalized to make them invariant under a scale change by using the normalized central moments instead of the central moments. The normalized central moments are defined by:
\[ M_{pq} = M_{pq}^a \quad \text{where} \quad a = \frac{(p + q)}{2} + 1 \]  

These, when substituted into the above equations, will give seven moments which are invariant to translation, scale change and rotation.

The \( \phi_i \) have large dynamic values. Thus, it was found that it was more practical to deal with the logarithms of the absolute value of the \( \phi_i \); thus; the seven moment invariants used in the proposed system are replaced by their logarithmic values. Table 1 shows the rounded values of \( \phi \) obtained for some of the words in the training set.

<table>
<thead>
<tr>
<th>( \Phi_1 )</th>
<th>( \Phi_2 )</th>
<th>( \Phi_3 )</th>
<th>( \Phi_4 )</th>
<th>( \Phi_5 )</th>
<th>( \Phi_6 )</th>
<th>( \Phi_7 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1510</td>
<td>0.9937</td>
<td>1.2045</td>
<td>0.3176</td>
<td>0.4587</td>
<td>1.5390</td>
<td>0.5082</td>
</tr>
</tbody>
</table>

**B. Zernike moments**

Zernike [12] defined a complete orthogonal set \( \{V_{nm}(x, y)\} \) of complex polynomials over the polar coordinate space inside a unit circle \( x^2 + y^2 = 1 \) as follows:

\[ V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{in\theta} \]  

Where \( j = \sqrt{-1}, \ n \geq 0, \ m \) is a positive or negative integer, \( |m| \leq n, \ n - |m| \) is even, \( \rho \) is the shortest distance from the origin to \( (x, y) \) pixel, \( \theta \) is the angle between vector \( \rho \) and \( x \)-axis in counter clockwise direction, and \( R_{nm}(\rho) \) is the orthogonal radial polynomial given by:

\[ R_{nm}(\rho) = \sum_{s=0}^{\frac{|m|}{2}} (-1)^s s! \left( \frac{n+|m|}{2} - s \right)! \left( \frac{n-|m|}{2} - s \right)! \rho^{n-2s} \]  

Note that \( R_{m-n}(\rho) = R_{nm}(\rho) \). These polynomials are orthogonal and satisfy the following condition:

\[ \int_{x^2 + y^2 \leq 1} V_{pq}(x, y) V_{pq}(x, y) \ dx dy = \frac{\pi}{n+1} \delta_{pq} \delta_{mq} \]  

Where:

\[ \delta_{ab} = \begin{cases} 1; & \text{if} \ a = b \\ 0; & \text{otherwise} \end{cases} \]  

Zernike moments are the projection of the image intensity function \( f(x, y) \) onto the complex conjugate of the previously defined Zernike polynomial \( V_{nm}(\rho, \theta) \), which is defined only over the unit circle:

\[ A_{nm} = \frac{n+1}{\pi} \int_{x^2 + y^2 \leq 1} f(x, y)V_{nm}(\rho, \theta)dx dy \]  

For a digital image, Zernike moments are given by:

\[ A_{nm} = \frac{n+1}{\pi} \sum_{x,y} f(x, y)V_{nm}^{*}(\rho, \theta), x^2 + y^2 \leq 1 \]  

Figure 2 shows an example of Zernike moments calculated using a sample word from the IFN/ENIT database.

**C. Pseudo-Zernike moments**

Pseudo-Zernike moments are derived from Zernike moments; therefore, both these moments have analogous properties [4]. The difference lies in the definition of the radial polynomial, which is given by:

\[ R_{nm}(\rho) = \sum_{s=0}^{\frac{|m|}{2}} (-1)^s s! \left( \frac{2n+1-s}{2} \right)! \left( \frac{n-|m|}{2} - s \right)! \rho^{n-2s} \]  

Where \( n \geq 0, \ m \) is a positive or negative integer subject to \( |m| \leq n \) only. Replacing the radial polynomial in Zernike moments with the radial polynomial defined above yields pseudo-Zernike moments. Pseudo-Zernike moments offer more feature vectors than Zernike moments due to the condition that \( n - |m| \) is even for the latter, which reduces the polynomial by almost half. In addition, pseudo-Zernike moments perform better on noisy images.
D. Tchebichef moments

In these moments the basis function is the discrete orthogonal Tchebichef polynomial. For a given positive integer \( N \) (normally the image size), the Tchebichef polynomial \([4]\) is given by the following recurrence relation:

\[
t_n(x) = \frac{(2n-1)t_{n-1}(x) - (n-1)(1 - \frac{(n-1)^2}{N^2})t_{n-2}(x)}{n}
\]

With the initial conditions:

\[
t_0(x) = 1, \quad t_1(x) = 2x + 1 - N/N
\]

Where \( n=0,1,...,N-1 \). The Tchebichef moment of order \((p+q)\) of an image intensity function is defined as:

\[
T_{nm} = \frac{1}{\rho(n,N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} t_m(x) t_n(y) f(x,y)
\]

Where \( n,m=0,1,...,N-1 \). The Tchebichef polynomial satisfies the property of orthogonality with:

\[
\rho(n,N) = \frac{N(1 - \frac{1}{N^2})(1 - \frac{2^2}{N^2})...(1 - \frac{n^2}{N^2})}{2n+1}
\]

Note that with Tchebichef moments, the problems related to continuous orthogonal moments are purged by using a discrete orthogonal basis function (i.e., Tchebichef polynomial) \([4]\). In addition, no mapping is required to compute Tchebichef moments as the Tchebichef polynomials are orthogonal in the image coordinate space.

\[
p_p(x) = \frac{1}{2^p p!} \frac{d^p}{dx^p} (x^2 - 1)^p, x \in [-1,1]
\]

\[
L_{pq} = \frac{(2p+1)(2q+1)}{4} \int_{-1}^{1} p_p(x)p_q(y)f(x,y)dx dy
\]

Where \( p, q \) are integers between \((0,\infty)\), and \( x \) and \( y \) are the \( x \)- and \( y \)-coordinate of the image.

Similarly, the Legendre moment for a \((N \times N)\) digital image is given by:

\[
L_{pq} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} p_p(m)p_q(n)f(m,n)
\]

Where:

\[
m_n = \frac{2m-N+1}{N-1}
\]

To compute the Legendre moments, a given image is mapped into the limit domain \([-1,1]\) as the Legendre polynomial is only defined over this range.

IV. Multi Layer Perceptron

MLP networks are widely used in handwritten recognition systems because they are very easy to train and very fast to use in classification decision processes. This popularity is related to the use of the gradient back-propagation algorithm \([7]\) in the training process. MLPs generally achieve good performance in terms of correct recognition rate in handwritten classification. When adopting MLPs, the word image is pre-processed so as to yield a feature vector, which is used for feeding the neural network. The class membership is commonly given by exclusive coding of the output (one-hot coding). During the recognition phase, the network is fed with the input pattern which is propagated through forward steps to the outputs. The expected class is simply given by the output unit with the highest value. The classifier can reject patterns whose membership cannot be clearly established.

A typical classification criterion which is used consists of rejecting a pattern if:

\[
\bar{y} = \max_{i=1,...,n} \{y_i\} < RM
\]

Where \( n \) is the number of classes, \( y_i \in (0,1) \) is the \( i \)-th output of the network, and \( RM \) is a proper threshold. Unfortunately, there are limits when using MLPs in classification tasks: First, there is no theoretic relationship between the MLP structure (ex: hidden layers number and nodes number per layer) and the classification task. The
second limitation is due to the fact that MLP derives separating hyperplane surfaces, in feature representation space, which are not optimal in terms of the margin area between the examples of two different classes.

To train the MLP network, we have used back propagation algorithm (BP) [6]. This algorithm performs a gradient descent in the connection weight space on an error surface defined by:

\[ E = \frac{1}{P} \sum_p E_p \]  

(29)

where:

\[ E_p = \frac{1}{2} \sum_k (t_{pk} - y_{pk})^2 \]  

(30)

Here P is the total number of patterns in the training set, and \( \{ t_{pk} \}, \{ y_{pk} \} \) are respectively the target and output vectors corresponding to the p-th input pattern. The quantity E is called the system error. In BP algorithm, weight updating rules are given by:

\[ w_{jk}(t+1) = w_{jk}(t) - \eta \frac{\partial E_p}{\partial w_{jk}}(t) + \alpha \Delta w_{jk}(t-1) \]  

(31)

\[ w_i(t+1) = w_i(t) - \eta \frac{\partial E_p}{\partial w_i}(t) + \alpha \Delta w_i(t-1) \]  

(32)

where \( w_{jk}(t) \) is the weight connecting a hidden node j with an output node k, while \( w_i(t) \) is the weight connecting an input node i with a hidden node j at time t. \( \Delta w_{jk}(t-1) \) is the modification amount to \( w_{jk} \) at time \( (t-1) \). \( \eta(>0) \) and \( \alpha(0<\alpha<1) \) are respectively called the learning rate and moment factor.

V. EXPERIMENTAL RESULTS

We tested our classifier: The Multi Layer Perceptron on the benchmark IFN/ENIT database of Arabic city names [9]. It was produced by the Institute for Communications Technology at the Technical University of Braunschweig (IFN) and the “Ecole Nationale d’Ingénieurs de Tunis”. The total number of binary images of handwritten Tunisian town/village names is 26459. Those names were written by 411 writers, and they were labeled according to 946 name classes.

For Zernike, Pseudo-Zernike, Legendre and Hu moments no normalization is needed. On the other hand for Tchebichef moments we have used a normalization procedure to obtain a 100x400 pixels images.

Table 2 shows the recognition rate given by pseudo-Zernike moments is the best which explain the efficiency of this type of moments, and then comes the Zernike moments with 87%, at the end we observe the low accuracy obtained with Legendre moments.

<table>
<thead>
<tr>
<th>Moments type</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hu moments</td>
<td>85%</td>
</tr>
<tr>
<td>Legendre moments</td>
<td>76%</td>
</tr>
<tr>
<td>Tchebichef moments</td>
<td>82%</td>
</tr>
<tr>
<td>Pseudo-Zernike moments</td>
<td>89%</td>
</tr>
<tr>
<td>Zernike moments</td>
<td>87%</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

Orthogonal and geometric moments have been proven to have superior feature representation capability and low information redundancy.

Performance of different moments for feature extraction in Arabic word recognition was studied in this paper. The IFN/ENIT database was used to perform this study. The moments used in this study were Hu moments, Zernike moments, Pseudo Zernike moments, Tchebichef moments, and Legendre moments. A two layer Perceptron neural network was used as the classifier in the recognition system. Pseudo Zernike moments contain useful no-redundant information, therefore lead to the best results.

REFERENCES
