On Pattern Classification Using Statistical Moments

Hamid Reza Boveiri
Faculty Member, Sama College, Islamic Azad University, Shushtar Branch
Member, Young Researchers Club, Islamic Azad University, Shushtar Branch
boveiri@ymail.com

Abstract

Selecting appropriate feature extraction method is absolutely one of the most important factors to archive high classification performance in pattern recognition systems. Among different feature extraction methods proposed for pattern recognition, statistical moments seem to be so promising. Whereas theoretical comparison of the moments is too complicated, in this paper, an experimental evaluation on four well known statistical moments namely Hu invariant moments, Affine invariant moments, Zernike moments, and Pseudo-Zernike moments is presented. Set of different experiments on a binary images dataset consisting of regular, translated, rotated, and scaled Persian printed numerical characters using a nearest neighbor rule classifier has been done and variety of interesting results have been presented. Finally, the results show that Pseudo-Zernike moments outperform the other introduced moments.

Keywords: *Feature extraction, optical character recognition, pattern recognition, statistical moments.*

1. Introduction

Pattern recognition is the scientific discipline whose goal is the classification/clustering of objects into a number of categories or classes. Depending on the application, these objects can be images, signal waveforms, or any type of measurements that need to be classified. The generic term pattern is used to refer these objects [1]. A typical pattern recognition system is a system composed of some processing steps (Fig. 1):

1. Preprocessing: In this step, different operation such as segmentation, binarization, noise removal, etc would be done to enhance input patterns.

2. Feature extraction: Unique features are extracted from the input patterns, which must discriminate the patterns individually.

3. Classification: In this step, the input pattern will be recognized with respect to the features extracted by previous step and then shall be assigned to appropriate class using a classifier.

Applications of Pattern recognition systems and techniques are numerous and cover a broad scope of activities like medicine, marketing, military, machine vision, robotics, remote sensing, and so on.

Whereas classifier can only see extracted features of input pattern and it has to make final classification with respect to the features, selecting appropriate feature extraction method is one of the most important factors to archive high classification performance in pattern recognition systems. However, the other steps in the system also need to be optimized. A

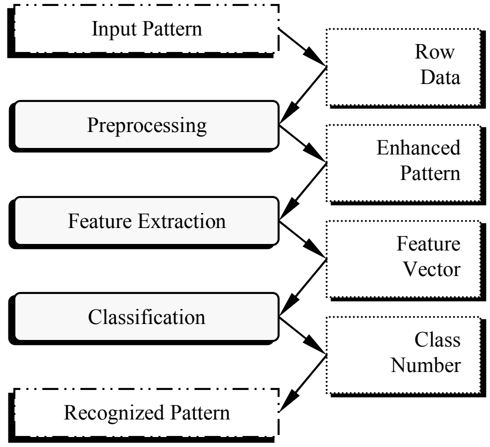

Figure 1. Structure of a typical pattern recognition system.

good feature extraction method is one that minimizes the within-class pattern variability while enhancing the between-class pattern variability [2]. However, the extracted features should be invariant to the expected distortions and variations that patterns may have in specific application. It should be clear that different feature extraction methods fulfill these requirements to a varying degree, depending on the specific recognition problem and available data. A feature extraction method that proves to be successful in one application domain may turn out not to be very useful in another domain.

Different feature extraction methods are divided into two categories structural and statistical. Among the statistical features, moments have achieved promising results [3]. Theoretical comparison of these moments is too complicated and maybe impossible. So experimental evaluation of the moments may be useful for practitioners and researchers gives them guidelines what kind of moments is more qualified to a specific application.

The rest of paper is organized as follow: next section studies theory of moments and introduces the four interested moments. Section 3 is devoted to implementation details such as the utilized classifier and dataset. Experimental results are presented in section 4 and finally, the conclusions are outlined in the last section.

2. Theory of Moments

Geometrical moment of order ($p+ q$) for a *two*-dimensional discrete function like image is computed by using (1). If the image can have nonzero values only in the finite part of xy plane; then moments of all orders exist for it [4].

$$m_{pq} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} x^p y^q f(x,y) \tag{1}$$

where $f(x, y)$ is image function and M, N are image dimensions. Then, by using (2) geometrical central moments of order equal to ($p+ q$) can be computed.

$$\mu_{pq} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{2}$$

where $\bar{x}$ and $\bar{y}$ are gravity center of image and are calculated by using (3). Actually by image translation to coordinate origin while computing central moments, they become translation invariant.

$$\bar{x} = \frac{m_{10}}{m_{00}} \ , \qquad \bar{y} = \frac{m_{01}}{m_{00}} \tag{3}$$

Note that in a binary image, $m_{00} = \mu_{00}$ is count of foreground pixels and has direct relation to image scale, therefore central moments can become scale normalized using (4).

$$\eta_{pq} = \frac{\mu_{pq}}{m_{00}^{a}} \quad , \quad a = \frac{p+q}{2} + 1 \tag{4}$$

2.1. Hu Invariant Moments

Moment invariants are a set of nonlinear functions, which are invariant to translation, scale, and orientation and are defined on normalized geometrical central moments. Hu [4] first introduced seven moment invariants based on normalized geometrical central moments up to the third order. Then, Li [5] extended the moments and listed 52 Hu invariant moments of order 2-9. Since the higher order moment invariants have resulted higher sensitivity, a set of twelve moment invariants limited by order less than or equal to four seems to be proper in most applications [6]. Having normalized geometrical central moments of order four and the lesser ones, seven moment invariants (φ_1-φ_7) introduced by Hu and then five extended ones (φ_8-φ_{12}) developed by Li, can be computed using (5) and (6) respectively.

$$\begin{aligned}
\varphi_1 &= \eta_{20} + \eta_{02} \\
\varphi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\varphi_3 &= (\eta_{30} - 3\eta_{12}) + (3\eta_{21} - \eta_{03})^2 \\
\varphi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\varphi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\varphi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + v_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
&\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\varphi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + 3(\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned} \tag{5}$$

$$\begin{aligned}
\varphi_8 &= \eta_{40} + \eta_{22} + \eta_{02} \\
\varphi_9 &= (\eta_{40} - \eta_{04})^2 + 4(\eta_{31} - \eta_{13})^2 \\
\varphi_{10} &= (\eta_{40} - 6\eta_{22} + \eta_{04})^2 + 16(\eta_{31} - \eta_{13})^2 \\
\varphi_{11} &= (\eta_{40} - 6\eta_{22} + \eta_{04})^2[(\eta_{40} - \eta_{04})^2 + 4(\eta_{31} - \eta_{13})^2] \\
&\quad + 16(\eta_{40} - \eta_{04}) + (\eta_{31} + \eta_{13})(\eta_{31} - \eta_{13}) \\
\varphi_{12} &= (\eta_{40} - 6\eta_{22} + \eta_{04})^2[(\eta_{40} - \eta_{04})^2 + 4(\eta_{31} - \eta_{13})^2] \\
&\quad - 16(\eta_{40} - \eta_{04}) + (\eta_{31} + \eta_{13})(\eta_{31} - \eta_{13})
\end{aligned} \tag{6}$$

2.2. Affine Moment Invariants

Flusser and Suk [7], [12] presented four affine moment invariants based on geometrical central moments up to the third order, which are invariant under general affine transformations. Whereas these four moments were not enough to discriminate patterns in most applications, in [8] they introduced a general method for systematic derivation of affine moment invariants of any weights and orders. They explicitly presented ten affine moment invariants can be computed by using (7).

$$
\begin{aligned}
I_1 &= (\mu_{20}\mu_{02} - \mu_{11}^2) / \mu_{00}^4 \\
I_2 &= (-\mu_{30}^2\mu_{03}^2 + 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} - 4\mu_{30}\mu_{12}^3 \\
&\quad -4\mu_{21}^3\mu_{03} + 3\mu_{21}^2\mu_{12}^2) / \mu_{00}^{10} \\
I_3 &= (\mu_{20}\mu_{21}\mu_{03} - \mu_{20}\mu_{12}^2 - \mu_{11}\mu_{30}\mu_{03} + \mu_{11}\mu_{21}\mu_{12} \\
&\quad +\mu_{02}\mu_{30}\mu_{12} - \mu_{02}\mu_{21}^2) / \mu_{00}^7 \\
I_4 &= (-\mu_{20}^3\mu_{03}^2 + 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 3\mu_{20}^2\mu_{02}\mu_{12}^2 \\
&\quad -6\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} - 6\mu_{20}\mu_{11}^2\mu_{12}^2 \\
&\quad +12\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 3\mu_{20}\mu_{02}^2\mu_{21}^2 \\
&\quad +2\mu_{11}^3\mu_{30}\mu_{03} + 6\mu_{11}^3\mu_{21}\mu_{12} - 6\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} \\
&\quad -6\mu_{11}^2\mu_{02}\mu_{21}^2 + 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} - \mu_{02}^3\mu_{30}^2) / \mu_{00}^{11} \\
I_5 &= (\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2) / \mu_{00}^6 \\
I_6 &= (\mu_{40}\mu_{22}\mu_{04} - \mu_{40}\mu_{13}^2 - \mu_{31}^2\mu_{04} + 2\mu_{31}\mu_{22}\mu_{13} - \mu_{22}^3) / \mu_{00}^9 \\
I_7 &= (\mu_{20}^2\mu_{04} - 4\mu_{20}\mu_{11}\mu_{13} + 2\mu_{20}\mu_{02}\mu_{22} + 4\mu_{11}^2\mu_{22} \\
&\quad -4\mu_{11}\mu_{02}\mu_{31} + \mu_{02}^2\mu_{40}) / \mu_{00}^7 \\
I_8 &= (\mu_{20}^2\mu_{22}\mu_{04} - \mu_{20}^2\mu_{13}^2 - 2\mu_{20}\mu_{11}\mu_{31}\mu_{04} \\
&\quad +2\mu_{20}\mu_{11}\mu_{22}\mu_{13} + \mu_{20}\mu_{02}\mu_{40}\mu_{04} \\
&\quad -2\mu_{20}\mu_{02}\mu_{31}\mu_{13} + \mu_{20}\mu_{02}\mu_{22}^2 + 4\mu_{11}^2\mu_{31}\mu_{13} \\
&\quad -4\mu_{11}^2\mu_{22}^2 - 2\mu_{11}\mu_{02}\mu_{40}\mu_{13} + 2\mu_{11}\mu_{02}\mu_{31}\mu_{22} \\
&\quad +\mu_{02}^2\mu_{40}\mu_{22} - \mu_{02}^2\mu_{31}^2) / \mu_{00}^{10} \\
I_9 &= (\mu_{30}^2\mu_{12}^2\mu_{04} - 2\mu_{30}^2\mu_{12}\mu_{03}\mu_{13} + \mu_{30}^2\mu_{03}^2\mu_{22} \\
&\quad -2\mu_{30}\mu_{21}^2\mu_{12}\mu_{04} + 2\mu_{30}\mu_{21}^2\mu_{03}\mu_{13} \\
&\quad +2\mu_{30}\mu_{21}\mu_{12}^2\mu_{13} - 2\mu_{30}\mu_{21}\mu_{03}^2\mu_{31} \\
&\quad -2\mu_{30}\mu_{12}^3\mu_{22} + 2\mu_{30}\mu_{12}^2\mu_{03}\mu_{31} + \mu_{21}^4\mu_{04} \\
&\quad -2\mu_{21}^3\mu_{12}\mu_{13} - 2\mu_{21}^3\mu_{03}\mu_{22} + 3\mu_{21}^2\mu_{12}^2\mu_{22} \\
&\quad +2\mu_{21}^2\mu_{12}\mu_{03}\mu_{31} + \mu_{21}^2\mu_{03}^2\mu_{40} - 2\mu_{21}\mu_{12}^3\mu_{31} \\
&\quad -2\mu_{21}\mu_{12}^2\mu_{03}\mu_{40} + \mu_{12}^4\mu_{40}) / \mu_{00}^{13} \\
I_{10} &= (-\mu_{50}^2\mu_{05}^2 + 10\mu_{50}\mu_{41}\mu_{14}\mu_{05} - 4\mu_{50}\mu_{32}\mu_{23}\mu_{05} \\
&\quad -16\mu_{50}\mu_{32}\mu_{14}^2 + 12\mu_{50}\mu_{23}^2\mu_{14} - 16\mu_{41}^2\mu_{23}\mu_{05} \\
&\quad -9\mu_{41}^2\mu_{14}^2 + 12\mu_{41}\mu_{32}^2\mu_{05} + 76\mu_{41}\mu_{32}\mu_{23}\mu_{14} \\
&\quad -48\mu_{41}\mu_{23}^3 - 48\mu_{32}^3\mu_{14} + 32\mu_{32}^2\mu_{23}^2) / \mu_{00}^{14}
\end{aligned} \tag{7}
$$

2.3. Zernike Moments

Zernike moments are a set of complex polynomials which form a complex orthogonal set over the interior of the unit circle, i.e. $x^2 + y^2 = 1$. The Zernike moments are projection of the input patterns on to the space spanned by the orthogonal $\{V_{nm}(x, y)\}$ functions [3]:

$$V_{nm}(x,y) = R_{nm}(x,y)\exp\left(jm\tan^{-1}\frac{y}{x}\right) \tag{8}$$

where n and m are order and repetition respectively and $j = \sqrt{-1},\ n \geq 0,\ |m| \leq n,\ n-|m|$ is even, and $R_{nm}(x, y)$ calculated as follow:

$$R_{nm}(x,y) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (x^2+y^2)^{(n/2)-s}(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} \tag{9}$$

For a *two*-dimensional pattern like image, the Zernike moment of order n and repetition m is given by:

$$A_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f(x,y)[V_{nm}(x,y)]^* \tag{10}$$

where $x^2 + y^2 \leq 1$ and the symbol $*$ denotes the complex conjugate operator. Note that the image coordination must be mapped to the range of unit circle. Khotanzad and Hong [9] used the amplitudes of the Zernike moments successfully in an offline optical character recognition, which are invariant under rotation but sensitive to translation and scale. Therefore, translation and scale normalization must be done using (11) before extraction of the Zernike moments.

$$f(x,y) = f\left(\bar{x}+\frac{x}{a}, \bar{y}+\frac{y}{a}\right) \tag{11}$$

where $(\bar{x}, \bar{y})$ being centroid of pattern function $f(x, y)$ and $a = (\beta / m_{00})^{1/2}$; β is a predetermined value for the number of object points in the pattern.

2.4. Pseudo-Zernike Moments

Difference between Pseudo-Zernike moments and Zernike ones described before is in moment repetition m. In Zernike moments, $n-|m|$ must be even while this rule does not exist in the Pseudo-Zernike ones. Therefore, number of the Pseudo-Zernike moments of order n is almost double number of the Zernike moments. New $R_{nm}(x, y)$ has some modifications, and should be computed by using (12) [10].

$$R_{nm}(x,y) = \sum_{s=0}^{n-|m|} \frac{(-1)^s (x^2+y^2)^{\frac{n-s}{2}}(2n+1-s)!}{s!(n-|m|-s)!(n+|m|+1-s)!} \tag{12}$$

Note that $V_{nm}(x, y) = V_{n,-m}(x, y)$ thus, only moments for positive m's need to be computed.

3. Implementation Details

The system was implemented on a Pentium IV (2.6GHz LGA) desktop computer with Microsoft Windows XP (SP2) platform using Microsoft Visual Basic 6.0 programming language. The utilized dataset and classifier will be discussed briefly as follow.

3.1. The Utilized Dataset

The dataset introduced in [11] has been used. This dataset consists of 64×64 binary images of all 10 Persian numeral characters (Fig. 2) in four groups. The first group is regular characters with same size and without any translations and rotations. The second group is rotated characters in the range of (-45, 45) degree. The third group is randomly translated characters and the characters of the fourth group have various sizes. There are 10 samples for each character in each group, in the other word; there are totally 400 samples (40 samples per character). Fig. 3 shows a sample image of each group of Persian character "three" in the dataset. In all experiments, half of samples have been used to train and the remainders as test data unless where it is expressed explicitly.

3.2. Nearest Neighbor Rule Classifier

A nearest neighbor rule classifier (NN) has been used to classify patterns in our implementation. In [6], it showed comparable results against multi layer Perceptron (MLP) and fuzzy min-max neural network (FMMNN). Nearest neighbor rule is a conventional nonparametric statistical classifier. In training stage, it stores all training samples in a table. Then in testing stage, it assigns an unknown input pattern to which class has minimum distance to a training sample of the class. Since it is a nonparametric classifier, classification will be done only based on nature of the introduced moments. Just such as the other classifiers, different definitions of distance can also be used here. While Euclidean distance has demonstrated high performance results, it was selected as distance measurement.

4. Experimental Results

Fig. 4 shows classification results of the patterns in the dataset with respect to using introduced moments as feature vector. Nearest neighbor rule, using Euclidean distance was utilized as pattern classifier. In all experiments, half of samples were used to train and the remainders as test data. Persian "seven" and "eight" are vertically reverse of each other. With respect to invariance of the moments to image reversal, they cannot discriminate these

Zero	One	Two	Three	Four
۰	۱	۲	۳	۴
Five	Six	Seven	Eight	Nine
۵	۶	۷	۸	۹

Figure 2. A typical Persian printed numerical characters.

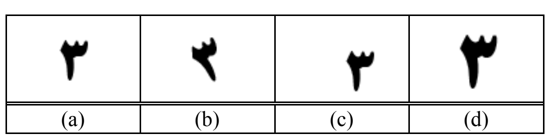

Figure 3. A sample image of each group of Persian "three" in the dataset. (a) Regular (b) Rotated (c) Translated (d) Scaled [11].

characters. Thus, character "eight" was dropped from the dataset in all experiments.

Fig. 4 (a) demonstrates correct classification diagram with respect to select different numbers of Hu invariant moments as feature vector. It is clear that increasing the number of utilized moments have not enhanced the correct classification rate. Whereas high orders of this moment show in high sensitivity, this behavior seems to be logical.

Fig. 4 (b) shows correct classification diagram with respect to select different numbers of affine moment invariants as feature vector. Despite Hu moment invariants that have sensitivity in higher order, affine moment invariant presents high sensitivity in certain moment numbers such as I_4, I_8, and I_9. Therefore, one can neglect them and use other moment numbers although these moments showed lower correct classification rate than three other moments and may not be proper for our application.

Fig. 4 (c) illustrates correct classification diagram with respect to select different orders of Zernike moment as feature vector. Numbers above the diagram are size of feature vector that consists of Zernike moments up to the given order. These moments showed objective behavior; correct classification rate is increased as the order of moments (and size of feature vector) increased. Despite of Hu and affine moment invariants, Zernike moments did not show high sensitivity even in high orders and their results were promising (100% correct classification rate was achieved by using feature vector consisting of Zernike moments up to

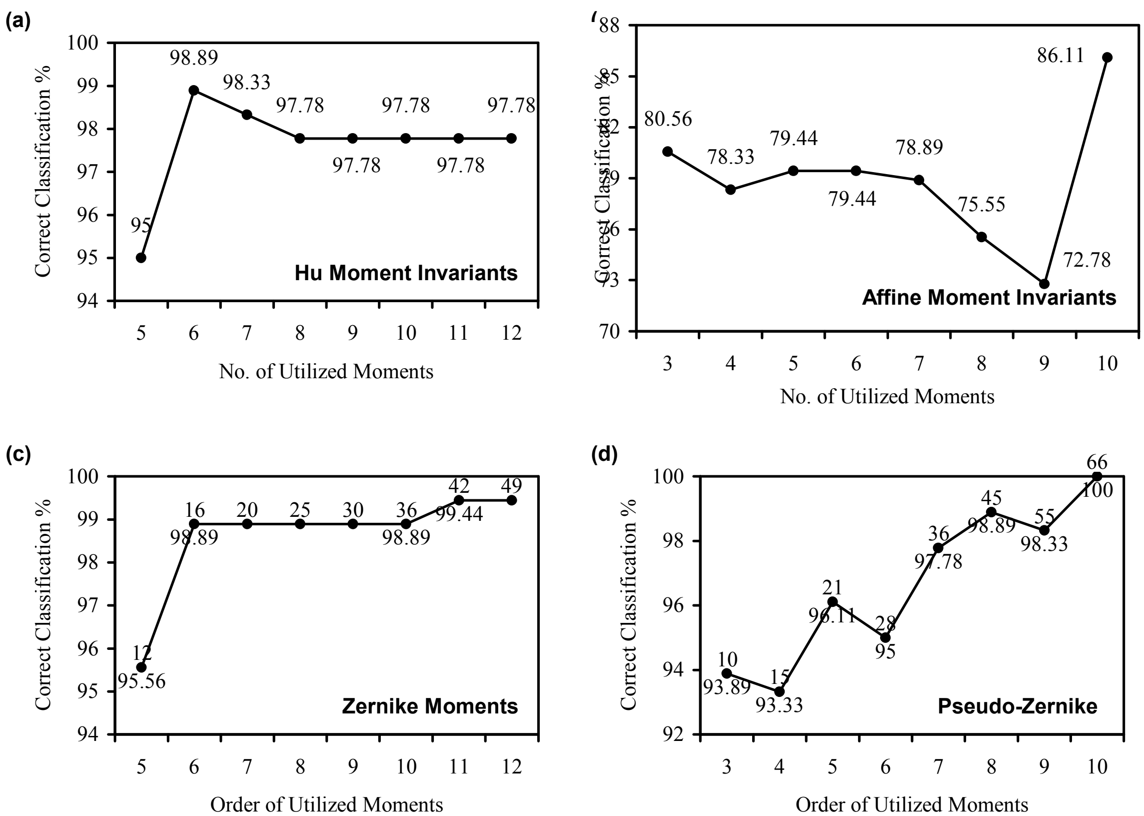

Figure 4. Classification results of the patterns in the dataset with respect to using four introduced moments as feature vector.

the order 13).

Fig. 4 (d) shows correct classification diagram with respect to select different orders of Pseudo-Zernike moment as feature vector. Numbers above the diagram are size of feature vector that consists of Pseudo-Zernike moments up to the given order. However, this moment shows sensitivity in some orders such as order 4, 6, and 9, eventually correct classification rate is increased as order of moments (and size of feature vector) increased. Since more Pseudo-Zernike features can be extracted in same order compared to Zernike ones, Pseudo-Zernike moments were achieved complete classification rate faster and finally, they outperformed the other introduced moments and seems to be more qualified.

5. Conclusion

In this paper, an experimental evaluation on pattern classification using statistical moments was introduced. Four well-known statistical moments namely Hu invariant moments, affine invariant moments, Zernike moments, and Pseudo-Zernike moments were considered as feature extraction methods. Set of different experiments on a binary images dataset consisting of regular, translated, rotated, and scaled Persian printed numerical characters using these moments and a nearest neighbor rule that is a conventional nonparametric statistical classifier was done and variety of results have been presented can be summarized as follow. Although correct classification rate should be enhanced generally while using more features (moments), there is some considerations. 1) Hu invariant moments showed some sensitivity in higher orders, so utilization of more moments will not enhance the results necessarily. 2) Some affine moment invariants such as $I4$, $I8$, and $I9$ showed high sensitivity. however, they had poor results, and seems to be unqualified in this case. 3) The achieved results using Zernike moments were promising. These moments had a little and monotonous sensitivity in all orders, so one can enhance the results by extracting more moments. 4) Even though, Pseudo-Zernike moments showed some sensitivity in some orders such as order 4, 6, and 9, finally they outperformed the other introduced moments.

Acknowledgment

This work has been supported by Sama College, Islamic Azad University, Shushtar Branch, Iran

References

[1] S. Teodoridis and K. Koutroumbas, *Pattern Recognition*, 2nd Ed., CA: Academic Press, 2003.

[2] P. A. Devijver and J. Kittler, *Pattern Recognition: A statistical Approach*, Prentice-Hall, London, 1982.

[3] O. D. Trier, A. K. Jain, T. Taxt, "Feature Extraction Methods for Character Recognition- A survey," *Pattern Recognition*, vol. 29, no. 4, pp. 641-662, 1996.

[4] M. K. Hu, "Visual Pattern Recognition by Moment Invariant," *IRE Trans. Info. Theory*, vol. IT – 8, pp. 179–187, Feb. 1962.

[5] Y. Li, "Reforming the Theory of Invariant Moments for Pattern Recognition," *Pattern Recognition Lett.*, vol. 25, pp. 723-730, Jul. 1992.

[6] H. R. Boveiri, "Persian Printed Numerals Classification Using Extended Moment Invariants," in *Proc. WASET Int. Conf. on Image and Vision Computing*, Rio de Janeiro, 2010, pp. 167-174.

[7] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition,* vol. 26, pp. 167-174, Jan. 1993.

[8] J. Flusser and T. Suk, "Graph Method for Generating Affine Moment Invariants," *IEEE*, 2004.

[9] A. Khotanzad and J-H. Lu, "Classification of Invariant Image Representations using a Neural Networks," *IEEE Trans. on Acoust., Speech and Signal Process.*, vol. 38, no. 6, pp. 1028-1038, Jun. 1990.

[10] C. H. Teh and R. T. Chin, "On Image Analysis by the methods of moments," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 496-513, 1988.

[11] H. R. Boveiri, "Persian Printed Numeral Characters Recognition Using Geometrical Central Moments and Fuzzy Min-Max Neural Network," *WASET International Journal of Signal Processing*, vol. 6, no. 2, pp. 76-82, 2010.

[12] J. Flusser and T. Suk, "Affine moment invariants: A new tool for character recognition," *Pattern Recognition Lett.*, vol. 15, pp. 433-436, Apr. 1994.

Hamid Reza Boveiri received the B.Sc. degree from Birjand University, Birjand, Iran, in 2005 and M.Sc. degree from Islamic Azad University Science and Research Branch, Ahwaz, Iran, in 2009, both in software engineering.

He is currently a faculty member of computer department of Sama College, Islamic Azad University Shushtar Branch, Shushtar, Iran. He is also a member of Young Researchers Club of Islamic Azad University Shushtar Branch and there he is advisor of research workgroup. His research interests include optimization, meta-heuristics, signal processing, and pattern recognition.