© 2013, Scienceline Publication
www.science-line.com
ISSN: 2322-5114
Journal of World's Electrical Engineering and Technology
J. World. Elect. Eng. Tech. 2(3): 47-53, 2013
JWEET

A Hybrid Method for Gesture Recognition

Mohsen Khalilpour and Navid Razmjooy*

Department of Electrical Engineering, Ardabil branch, Islamic Azad University, Ardabil, Iran

*Corresponding author's Email: navid.razmjooy@hotmail.com

Abstract *– In this article, we propose a new hybrid approach with Fourier technique well known as Fourier descriptor and HU moments to detect hand gesture signatures for vision-based applications. Fourier descriptors are efficient specifications of image feature extraction. They can be used for shape characterization by preventing the scale, translation, and rotation invariance, as well as their independent from small shape deviations. To obtain good results for hand images, they converted into HSV color space. HSV unlike RGB color space has some features help users understand the nuances in the images; since, if one color be similar to another color, HSV can get better results. After that, Hand image will be extracted from HSV image by implementing the threshold operation. The Fourier Descriptors of extracted hand images are hybridized with HU moments and are compared with the reference dictionary to perform gesture recognition. Experiments are performed on the Bochum Gestures database. Finally the performance of the gesture recognition using Euclidean distance investigated. Results show that proposed hybrid method gives a high performance for hand gesture applications.*

Keywords*: Hand gesture, HSV Color space, Hybrid Fourier Descriptor, Geometric Invariant Moments.*

ORIGINAL ARTICLE
Received 19 Oct. 2013
Accepted 27 Dec. 2013

INTRODUCTION

Vision-based hand gesture recognition methods include static hand gesture and their current locations with no movements involve them which can be utilized as an intuitive interaction between human and computer [1]. They have many advantages rather than devices like mice, keyboards or electronic gloves; especially when they are used directly to interact or communicate with a computer, thus they get more intuitive means of interplay [2]. To achieve this purpose, computers should be capable to detect hand gestures from the input image visually. There are already a large number of computational approaches and algorithms for the image segmentation tasks [3-7].

Lee devised a model of hand-gesture recognition by using Hidden Markov Model (HMM) [8]. Kjeldsen and Kendersi worked on skin-tone segmentation on HSV space according to the premise that skin tone in images indwells a connected volume in HSV space [9]. They also expanded a back-propagation neural network to detect gestures from the segmented hand images. Etsuko Ueda and Yoshio Matsumoto proposed a hand-gesture estimation technique which can be used for vision-based human interfaces; in the proposed approach, the hand regions are extracted from multiple images achieved by a multi-view point camera system, and forming the "pixel Model." Incorporating these multi-viewpoint images make hand gesture to reconstruct as a "pixel model". After that, all joint angles are estimated utilizing three dimensional models coordinating by hand model and pixel model [10]. Chan Wah Ng and Surendra Ranganath considered a hand gesture recognition system; they utilized image Fourier Descriptors (FDs) to represent the shape of hand blobs achieved from the input image as their prime feature; after that they classified the features by using a Radial Basis Function (RBF) neural network for the final gesture recognition [11]. Claudia Nölker and Helge Ritter described a hand gesturee recognition modal based on neural networks [12]. They utilized finger tips to recognition purpose; in their method full identification of all finger joint angles based on that a 3D modal are found. Their model developed according to the dimensions and movement possibilities of a natural human hand [12].

Some other work also has been implemented based on detecting hands in gray level images according to their appearance and texture. Wu and Huang proposed another hand posing technique by adding a large unlabeled training set; they combined the supervised and unsupervised learning to generate a powerful learning approach. The Discriminant-EM (D-EM) algorithm used as the synthetic approach; D-EM not only approximates the parameters of a generative model but also finds a linear transformation to relax the assumption of probabilistic structure of data distributions as well as choose good features automatically [13].

Rosales et al. proposed a non-linear supervised learning approach to estimate hand gestures from a single image; the specialized mappings architecture (SMA) is employed to map image features to likely 3D hand gestures. Although the described method assumes that the hand silhouette is correctly recognized in the input image, whilst such accurate hand detection is often unreal to presume in real world applications [14]. T. E. de Campos and D. W. Murray presented a regression based method for hand posing; the described approach included RVM-based learning method for estimating the whole body gesture which is adapted to hand gestures recovery [15].

SYSTEM OVERVIEW

This article presents a hybrid method for hand gesture recognition. In the first step, for hand detection purpose, the input image is converted from RGB color space into Hue color space image. After that, by using a definite thresholding and utilizing mathematical morphology on the images, the hand is detected. Afterwards, Fourier descriptors (FD) are used to hand feature extraction and posing. FDs are specific features derived from frequency space which are efficient for pattern recognition purposes. They used contour images for feature extraction.

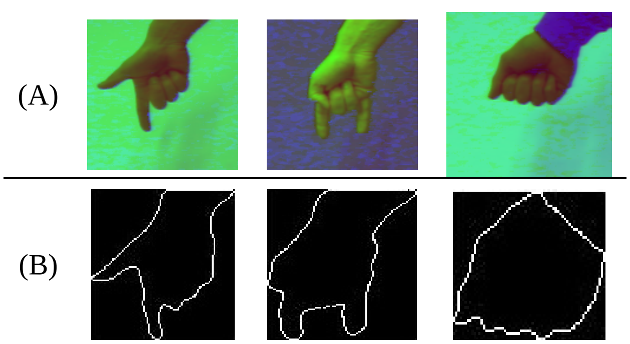


Fig.1 - (A) Original hand (HSI color space); (B) hand contour. Note that the number of descriptors is 40

FDs have already been examined for hand gesture recognition, but the existent approaches used the raw extracted data from Fourier Descriptors as the main features for detecting the hand gestures [16-19]. In this research we present a new combinational algorithm of FDs and geometric moments as the main features for hand gesture recognition. Block diagram of hand gesture recognition is shown below:

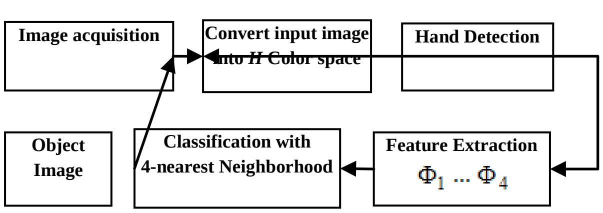


Fig. 2 - Block diagram of hand gesture recognition system

MATERIAL AND AMETHODS

A. Hand Segmentation

The first step in the hand posing is to detect the hand areas from background in order to get better results for the next step (i.e. feature extraction). Human skin detection can get much easier if we use HSV (hue-saturation-value) color space instead of RGB color space [20].

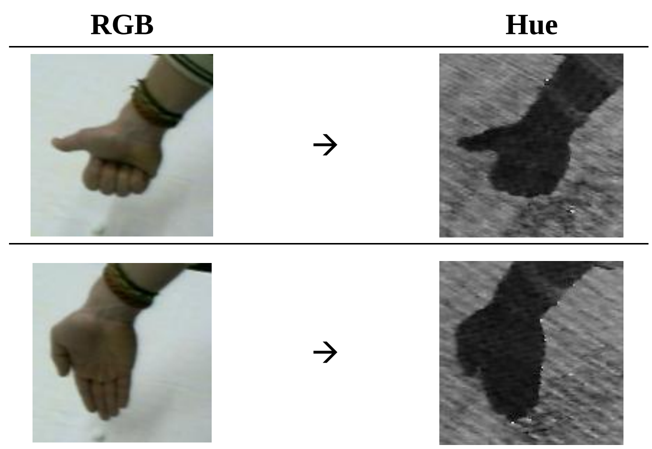


Fig. 3 - Two samples of RGB to H conversion

HSV is a most popular cylindrical-ordinates model of RGB which rearrange the geometry of RGB to get more perceptually relevant rather than the Cartesian representation. Experiments showed that Hue in HSV can give the proper result of hand detection purpose. The relationship among RGB and H color is shown below:

$$H' = \begin{cases} undifined, & if \quad C = 0 \\ \dfrac{G-B}{C} \bmod 6, & if \quad M = R \\ \dfrac{B-R}{C} + 2, & if \quad M = G \\ \dfrac{R-G}{C} + 4, & if \quad M = B \end{cases} \tag{1}$$

$$H = 60^{\circ} \times H' \tag{2}$$

Where M is the maximum value of R, G and B and C is the difference between M and the minimum value of R, G and B. For this reason, we use Hue of HSV model for hand segmentation. After image conversion the input image from RGB color space into the Hue color space, it is thresholded in a definite interval [21]. Of course, it is noticeable that threshold interval can be changed due to light radiance changes. Fig. 4 shows an estimation of 126 samples of hand feature vectors on Hue model.

Due to the color inconsistency of Hue model, the result of thresholding gets some misclassification. To overcome the presented problem, three morphological operations including: region filling, closing and area opening are utilized.

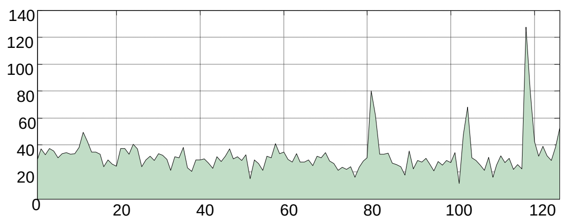


Fig.4 - A graphic model of Hue value interval for hand feature vectors

The algorithms for region filling are based on set dilation, complementation, and intersections, are illustrated below:

$$X_k = (X_{k-1} \oplus B) \cap A^C, \quad k = 1,2,3,\ldots \tag{3}$$

Where A is a set of boundary, B is the structure element, $X_{k-1} \oplus B$ is the dilation of X_{k-1} and B, $\cap$ is the sign unity and A^C is the complement of A. The algorithms terminate at iteration step k if:

$$X_k = X_{k-1}.$$

Area closing generally makes counters smooth, eliminates small holes, fuses narrow breaks, and long thin gulfs, and fills gaps in the contour. The closing of set A by structuring element B denotes as $A \bullet B$ and can be achieved as below [28]:

$$A \bullet B = (A \oplus B) \Theta B \tag{4}$$

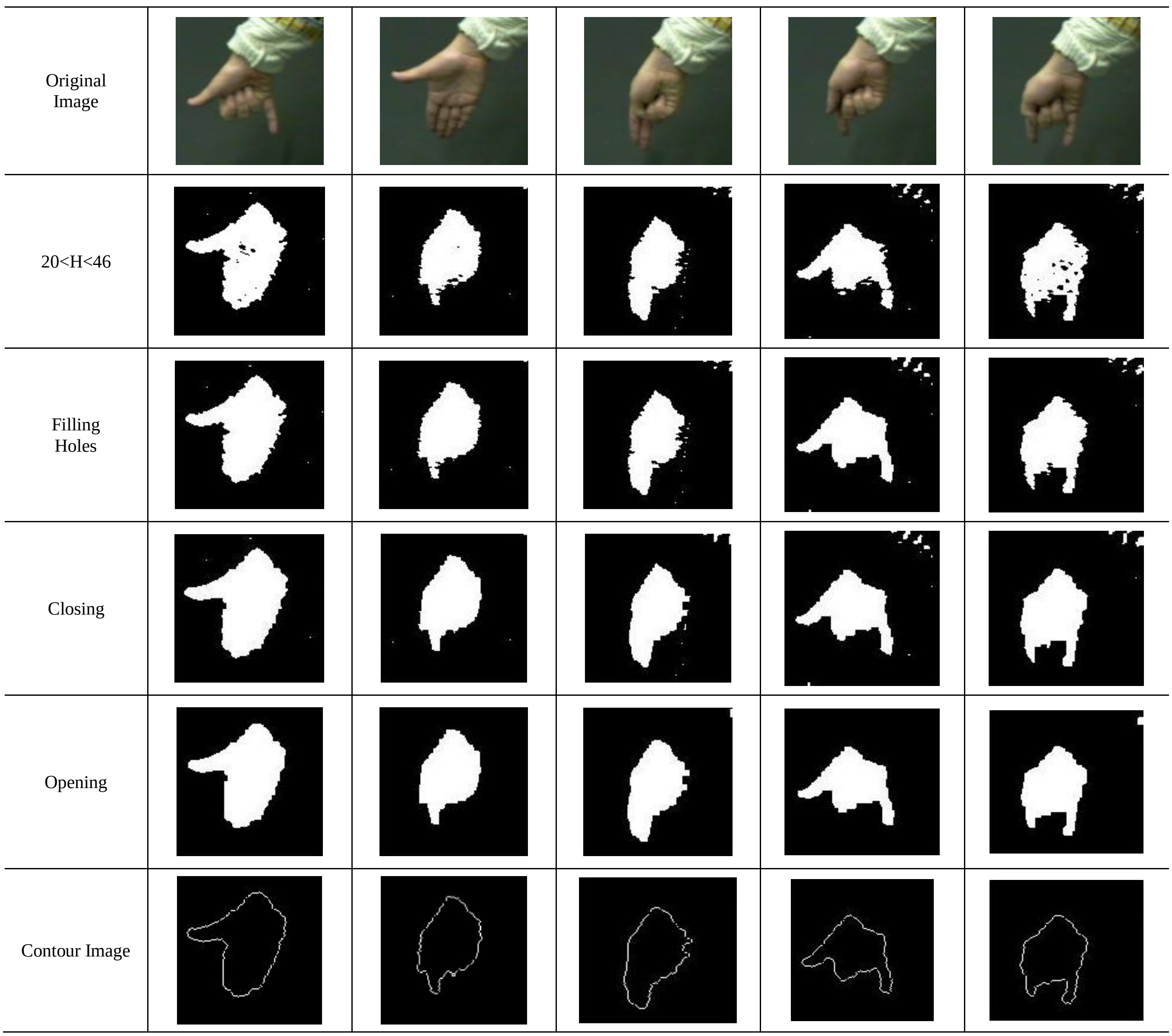


Fig. 5 - Five samples of hand segmentation process.

where the sign Θ denotes the erosion of two functions. Closing might result in synthesis of separated components, consequently, generating new holes.

Finally, the opening of *A* by *B* can be achieved by the erosion of *A* by *B*, followed by dilation of the resulting image by *B*:

$$A \circ B = (A \Theta B) \oplus B \quad (5)$$

Main idea of area opening is to eliminate small extra areas which are independent to hand area. More, 5 examples of the whole segmentation process are shown in fig.5.

B. Hand Gesture Recognition

Fourier descriptors (FDs) are considered to recognize the hand gesture. FDs are independent from rotation, translation and scale representations. There are some researches about using FDs as Features in hand gesture recognition [22, 23].

In this research, unlike other researches, we utilize Fourier descriptors just for estimating and achieving the main shapes signature. After that, geometric moments are used as features of the estimated images for the final gesture extraction. The main principle is based on that if similar shapes get estimated by some fewer features, the similarity of the homologous gestures can be detected more easily. It is noticeable that estimation point of hand region by FDs might be applying with proper numbers of estimation points [24].

Each one of the hand detected image is decomposed into its connected components which can be defined by their counters. FDs are applied on the contour of the pre-processed images to extract their componential features. The contour area can be demonstrated with different signatures like complex coordinates, curvature, central distance, cumulative angular function [25]. In this article the complex coordinates type is utilized. Each point M_i of the shape contour is demonstrated by a complex number z_i, with *N* the number of points of the contour:

$$\forall i \in [1, N], M_i(x_i, y_i) \Leftrightarrow z_i = x_i + jy_i \quad (6)$$

Fig. 6 shows that the low frequency coefficients which include information on the common form of the shape and the high frequency coefficients which include information on the small details of the shape. We can notice that with more than 40 coefficients the hand shape is reconstructed for the defined purpose.

Geometric invariant moments (GM) are then used to feature extraction of the reconstructed images [26]. GM features are usually used to extract a rotation, scale and Translation invariant features. Two-dimensional moments of a digital image $f(x, y)$ are presented below:

$$m_{pq} = \sum_x \sum_y x^p \times y^q \times f(x, y) \quad (7)$$
$$p, q = 1, 2, 3, \ldots, \infty$$

where *x, y* are the components of a two-dimensional image and *p+q* define the order of the moment.

The main idea in this paper is to hybridize two strong descriptors including Fourier Descriptors and Geometric invariant moments to achieve better performances toward single-mode methods.

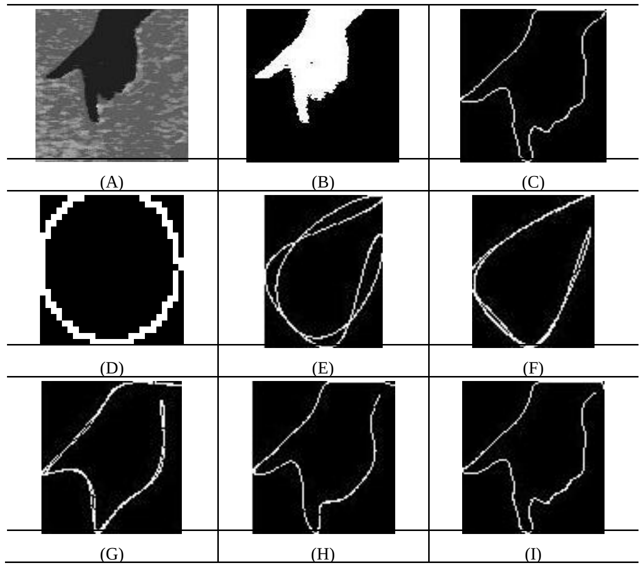


Fig.6. Example of reconstruction with FDs, as a function of the cut-off frequency, with an initial contour sampled at (D)2, (E) 8, (F) 10, (G) 20, (H) 40 and (I) 100 points. (A), (B) and (C) are the Hue model of the image, threshold image and contour shaped image respectively.

The central moments of $f(x, y)$ are defined as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \times (y - \bar{y})^q \times f(x, y) \quad (8)$$

where $(\bar{x}, \bar{y})$ are the centroid of the moments.

In the above formula, $\bar{x}$ and $\bar{y}$ are the coordinates of the centre and are as below:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad (9)$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \quad (10)$$

The normalized central moment of order (p+q) can be demonstrated as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{pq}^{\gamma}} \quad (11)$$

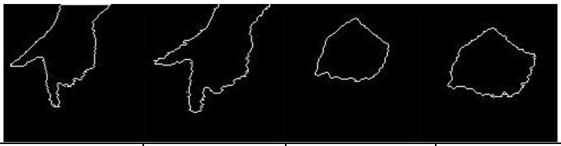
$$\gamma = \frac{p+q}{2} + 1 \quad (12)$$

Four functions of GMs are calculated from central moments through order three [27]:

$$\Phi_1 = \eta_{20} + \eta_{02}$$
$$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$\Phi_4 = (\eta_{30} - \eta_{12})^2 + (\eta_{21} - \eta_{03})^2$$

(13)

More, two examples of different images are demonstrated. Fig. 7 shows the example images used for the feature extraction purpose:



(A1) | (A2) | (B1) | (B2)

Fig.7 - Example images used for feature extraction.

Table 1 shows the central moments of the images above:

Table 1
Geometric invariant moments for 4 examples

Image	Φ_1	Φ_2	Φ_3	Φ_4
A1	0.208	0.043	0.001	0.001
A2	0.209	0.044	0.000	0.000
B1	0.144	0.021	0.000	0.001
B2	0.154	0.024	0.000	0.001

To improve the viewing area, we figure the table on the line chart (Fig. 8). As it can be seen from the above chart, A1 and A2 have almost similar values of central moments; the case is also correct for B1 and B2.

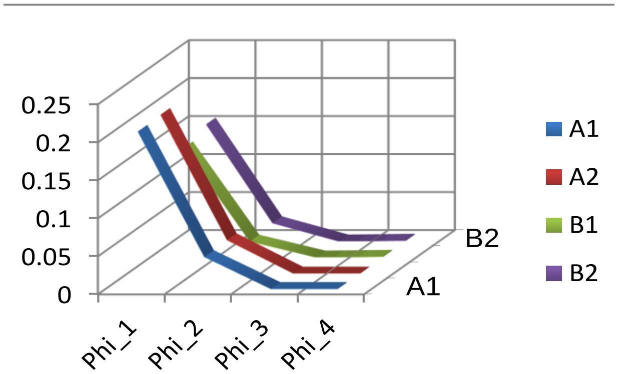


Fig. 8. Line chart of the four example images after feature extraction.

C. Classification

Hand shapes are classified with using a K-nearest-neighborhood. The K-nearest-neighborhood (Knn) algorithm is a generic and nonparametric supervised pattern classifier. In our application, Knn considers the distance between a query feature vectors and a set of vectors in the data set. The distance between two vectors can be computed using some distance function $d(x, y)$, where x and y are vectors created of N features, such that $x=\{x_1, x_N\}$, $y= \{y_1 \ldots y_N\}$. In this work different K values (K=3, 4, 5, 6) are utilized for nearest neighborhood [28, 29].

Euclidean distance is used in this work as the Knn metric:

$$d_E(x,y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \tag{14}$$

According to the distance between two vectors is related to the distance, the achieved distances are scaled with the arithmetic mean across the dataset is 0 and the standard deviation 1. This can be performed by replacing the scalars x and y with x' and y' using the following function:

$$x' = \frac{x - \bar{x}}{\sigma(x)} \tag{15}$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{16}$$

$$\delta(x) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{17}$$

where x is un-scaled value, $\bar{x}$ is the arithmetic mean of character x across the data set is its standard deviation, and is the achieved scaled value [21].

RESULTS AND DISCUSSION

Data base of static hand gestures collected from Bochum Gestures Database [30]. All images in database are 128×128 tiff images. The whole images were in two types of RGB and HSI which all converted to Hue model for the considered purpose. 100 images are selected for the train and test in this work and the classes which are used for classifications are 5. As it can be illustrated, numbers of features of the images for the feature extraction are also 4.

The performance is evaluated by using three quantitative quality criterions include CDR, FAR and FRR;

The first criterion is the correct detection rate (CDR) which is given in Eq.18. The false acceptance rate (FAR) is the percentage of recognition moments in which false acceptance occurs (Eq.19). The false rejection criterion (FRR) is the percentage of recognition moments in which false rejection happens (Eq.20).

$$CDR = \frac{\text{Number of data Correctly Classified}}{\text{Total data in the Test Dataset}} \tag{18}$$

$$FAR = \frac{\text{No. of unrelated hand poses Classified as related poses}}{\text{Total Pixels in the Test Dataset}} \tag{19}$$

$$FRR = \frac{\text{No. of related poses Classified as unrelated poses}}{\text{Total Pixels in the Test Dataset}} \tag{20}$$

Table 2 presents the performance of the proposed method with different values of K and Table 3 illustrates a comparison among the hybrid proposed method and two single-mode methods; note that the achieved values are the mean value of five gestures of: Turn right, Turn left, Forward, Stop and Backward.

Table 2
Classification Results of Performance in the presented procedures with different *K*.

Accuracy Metric	Classification Rate (%)			
K Value	K=3	K=4	K=5	K=6
CDR	94.0	94.0	95.0	96.0
FAR	04.0	03.0	03.0	03.0
FRR	02.0	03.0	02.0	01.0

Table 3
Comparison of The Proposed Method with Two different Techniques.

Accuracy Metric	Proposed Method K=6	Classification with FD [31]	Classification with Hu moments [32]
CDR	96.0	94.0	93.0
FAR	03.0	04.0	04.0
FRR	01.0	02.0	03.0

CONCLUSION

This article has presented a high performance approach for hand gesture recognition in the context of human computer interaction. In this research, a simple and efficient method is used to image segmentation. After that, a hybrid method including Fourier Descriptors and Hu moments is utilized for extracting the hand shape features. Finally, we have discussed a three nearest neighbor classifier to classify the images into relevant and un-relevant of gestures.

REFERENCES

[1] V. Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," IEEE trans. on Patt. Analysis and Mach. Intell. 19, 1997, pp. 677-692.

[2] S. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," IEEE trans. on Patt. Analysis and Mach. Intell. 27, 2005, pp. 873-891.

[3] E. Kukula and S. Elliott., "Implementation of hand geometry: An analysis of user perspectives and system performance," IEEE Aerospace Elec. Sys. Magazine, 2 IEEE trans. on Patt. Analysis and Mach. Intell. 2006, pp. 3-9.

[4] E. Konukog˘lu, E. Yörük, J. Darbon, and B. Sankur, "Shape-based hand recognition," IEEE Trans. on Image Proc., 15, 2006 pp. 1803-1815.

[5] E. Yörük, H. Dutagaci, and B. Sankur, "Hand biometrics, Image Vis. Computations," 24, 2006, pp. 483-497.

[6] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, "Biometric identification through hand geometry measurements," IEEE Trans. on Patt. Analysis, 22, 2000, pp. 1168-1171.

[7] Y. Bulatov, S. Jambawalikar, P. Kumar, and S. Sethia, "Hand recognition using geometric classifiers. Proc. DIMACS Workshop on Comp," Geometry, 2002, pp. 14–15.

[8] Hyeon-Kyu Lee and Jin H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition," IEEE trans. on pattern analysis and machine intell., 21, 1999.

[9] Rick Kjeldsen and John Kender, "Finding skin in color images," Proc. IEEE Int. Conf. on autom. Face and Gesture Recog., 1996, pp. 312-317.

[10] Etsuko Ueda, Yoshio Matsumoto, "Masakazu Imai, Tsukasa Ogasawara," Hand Pose Estimation for Vision-based Human Interface. IEEE Trans. on Indu. Elec., 50, 2003, pp. 676-684.

[11] Chan Wah Ng, Surendra Ranganath, "Real-time gesture recognition system and application," Image Vision Comp., 20, 2002, pp. 993-1007.

[12] Claudia Nölker and Helge Ritter, "Visual Recognition of Continuous Hand Postures," IEEE trans. on sys, man, and cybernetics—part c: applications and reviews, 31, 2001.

[13] Y. Wu and T. S. Huang, "View-independent Recognition of Hand Postures," Proc. IEEE Conf. on Comp. Vision and Patt. Recog., 2, 2000, PP. 84-94.

[14] R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff, "3D hand pose reconstruction using specialized mappings," Proc. ICCV, 2001, PP. 378–385.

[15] T. E. de Campos and D. W. Murray, "Regression-based hand pose estimation from multiple cameras," Proc. IEEE Conf. on Comp, Vision and Patt. Recog, 1, 2006, PP. 782–789.

[16] A. Kumar and D. Zhang, "Personal recognition using hand shape and texture. IEEE Trans. on Image Proc., 15, 2006, PP. 2454-2461.

[17] W. Li, D. Zhang, and Z. Xu, "Palm print identification by Fourier transform," Int. J. of Patt. Recog. Artificial Intell., 16, 2001, PP. 417-432.

[18] A. Kumar and H. C. Shen. " Palm print identification using Palm-Codes," Proc. 3rd Int. Conf. Image & Graphics, 2004, PP. 258–261.

[19] A. Kumar and D. Zhang, "Personal authentication using multiple palm print representation," Patt. Recog., 38, PP. 1695-1704.

[20] N. Tanibata, N. Shimada, and Y. Shirai, "Extraction of hand features for recognition of sign language words," Proc. Int. Conf. Vision Interface, 2002, PP. 391-398.

[21] J. Triesch and C. von der Malsburg, "A System for Person-Independent Hand Posture Recognition against Complex Backgrounds," IEEE PAMI, 23, 2001, PP. 1449-1453.

[22] S. Conseil, S. Bourennane and L. Martin, "Comparison of Fourier descriptors and HU moments for hand posture recognition," Proc. 15th European Sig. Process. Conf., 2007, PP. 1960-1964.

[23] Hardy Francke, Javier Ruiz-del-Solar and Rodrigo Verschae, "Real-time Hand Gesture Detection and Recognition using Boosted Classifiers and Active Learning," Proc. of 2nd Pacific Rim conf. on Advances in image and video tech., 2007, 2007, PP. 533-547.

[24] Thomas R. Crimmins, "A Complete Set of Fourier Descriptors for Two-Dimensional Shapes," IEEE Trans. Syst. Man and Cybernetics 6, 1982, pp.848-855.

[25] D. Zhang and G. Lu., "A comparative study on shape retrieval using Fourier descriptors with different shape signatures," Proc.Int. Conf. on Intell. Multimedia and Dist. Edu., 2001.

[26] S. A. Dudani, K. J. Breeding, and R. B. McGhee, "Aircraft identification by moment invariants," IEEE Trans. Computers, vol. 26, pp. 39–45,1977.

[27] M. K. Hu, "Visual patt. Recognition by moment invariants," IRE Trans. Inf. Theory, 8, 1962, PP. 179-187.

[28] Lei Wang, Latifur Khan and Bhavani Thuraisingham, “An Effective Evidence Theory based K-nearest Neighbor (KNN) classification,” IEE Proc. Web Intell. And Agent tech. 2008, pp.798-801.

[29] N. Razmjooy, B.S. Mousavi, F. Soleymani,“ A real-time mathematical computer method for potato inspection using machine vision,” Comp. & Math. with Apps., 63, 2012, PP. 268-279.

[30] Data base of static hand gestures, collected at the Institut fuer Neuroinformatik, Ruhr-Universitaet Bochum, Germany.

[31] Harding, Peter RG, and T. Ellis. "Recognizing hand gesture using Fourier descriptors." Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Vol. 3. IEEE, 2004.

[32] Liu, Yun, Yanmin Yin, and Shujun Zhang. "Hand Gesture Recognition Based on HU Moments in Interaction of Virtual Reality." Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on. Vol. 1. IEEE, 2012.