HSV Brightness Factor Matching for Gesture Recognition System

Mokhtar M.Hasan mmwaeli@gmail.com
Department of Computer Science
Faculty of Science
BHU-Varanasi
Uttar Pradesh-221005, India

Pramod K. Mishra pkmisra@gmail.com
Department of Computer Science
Faculty of Science
BHU-Varanasi
Uttar Pradesh-221005, India

Abstract

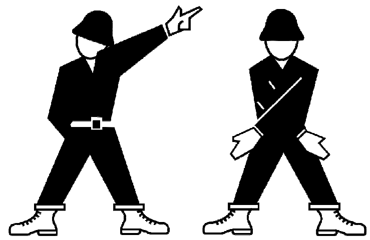
The main and primary objective of the gesture recognition research is to establish a system which can identify specific human gestures and utilize these identified gestures to be carried out by the human made machines, In this paper, we are going to introduce a new method for gesture recognition that based on the local brightness of each block of the gesture image, the input gesture image is divided into 25x25 blocks each of 5x5 block size, and we have calculated the local brightness of each divided block after applying colored segmentation operation using HSV (Hue, Saturation and Value) color model, so, each gesture produces 25x25 feature values called feature vector, our experimental shows that more that 65 % of these features are zero value which leads to minimum storage space, the recognition rate achieved is 91 % using 36 training gestures and 24 different testing gestures. We build a gesture recognition system that can communicate with the machine in natural way without any mechanical devices and without using the normal input devices which are the keyboard and mouse and the mathematical equations is the only translator that existed between the human and human-made machines, we have focused in this study on the hand gesture since hand can represents more meaning than other human-bodily organs.

Keyword: Brightness Calculation, HSV color model, Gesture Recognition, Template Matching, Image Segmentation, Laplacian Edge Detection.

1. INTRODUCTION

In all of the simulation processes, we tried to simulate the human abilities, in gesture recognition system, the remarkable ability of the human vision is the gesture recognition, it is noticeable mainly in deaf people when they communicating with each other via sign language and with hearing people as well. In this paper we tried to simulate this ability but this time will be between the human and human-made machines.

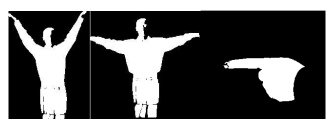
A gesture pose is a mode of communication between people that depends on the bodily movement especially the hand motion and pose; this form of communication is established along with spoken words in order to create a comprehensive statement to be carry out be the hearer. Most people use gestures language represented by bodily movement in addition to spoken words when they communicate between each other [1]; Figure (1) shows a gesture example for helicopter signaler.



A, take off B, landing

FIGURE 1: Helicopter Signaller for Marshaling Operations [2].

The normal communication between the people is the speaking which needs the sound to convey the meaning, while the later kind needs the space to convey the meaning [3]. The coarse classification of gestures is two; static and dynamic, the static gesture is a specific hand pose formed by a single image. The dynamic gesture is a moving gesture formed by a sequence of images [3] as in Figure (2),



A, dynamic B, static

FIGURE 2: A and B Represent the Dynamic and Static Gesture Respectively.

The application of gesture system on interactive applications produces many challenges. The first and important challenge is the response time which should be fast [4]. There should be no noticeable time between user gesture movement and computer replies [4]. The designed computer vision algorithms should be reliable and work for different ethnic people [4] especially when the color of human is changed comparing with white and black people. One more challenge which is the cost challenge, the gesture system needs special hardware such as the camera and sensors as necessarly, those special hardware will be the replacement of the existing hardware devises which may considered as low cost [4] such as the keyboard and mouse, but the gesture system with these new devices will be more worthwhile for wire-less communication.

This paper applied a new gesture recognition method for identifying the gestures for the computer or for the telerobotic in order to understand and carry on the human teleoperations, we have applied this novel method by windowing the image in order to recognize the input gesture and discover the meaning for that gesture. We have applied the proposed method using six gestures database, each of ten samples, so the total is sixty gestures database used for gesture recognition; we used the hand gesture rather than the face because the hand is the most flexible part of the body and can shows different meaning.

2. RELATED WORK

Roberto and Tomaso [5] applied a face recognition template matching; a very simple and direct recognition technique based on the use of whole image as grey-level templates. The most direct of the matching procedures is correlation. First, the image is normalized to obtain aunified locations for the mouth, eye and nose. The authors in [5] applied her technique by creating a database entry for each person contains the frontal view of that person, along with four masks which are the eyes, nose, mouth, and face (the region from eyebrows downwards as decided by the authors in [5]). All these four masks are relatively to the position of normalized eye position in whole of their database. The recognition measure applied by them is the Euclidian Distance; by matching the new presented face gesture with all the database gestures, and the database gesture corresponds to the maximum matching score is the recognized gesture, they had used samples taken from 47 persons with 4 gestures each.

Freeman and Roth [6] applied hand gesture recognition using orientation histogram, They had applied some transformation T to the image data in order to create the feature vector that will be used for recognition purpose and represents that specific gesture. To classify the gesture, they compare the feature vector with the feature vectors from a previously generated training set. The transformation T can be described as a polar representation for the histogram of the local orientations of the input gesture, they use the gradient direction to calculate this orientation, the histogram of the directions is then sketched using polar plot which represents the final features of the input gesture and the other features will treated the same, they had used samples taken from one person with 5-15 gestures.

K. Symeonidis in [4] applied gesture recognition method using neural network, he has used 8 different hand gestures each of 3 samples, those were for training purpose, he did not use an exact number of gestures for testing purpose since some gestures tolerate more samples than others, the features than he has used were the 19 elements degree numbers that been converted from the polar representation of the orientation histogram of the input gesture, then he presents these features for training the neural network after some preprocessing that casted the original features into later number of features.

Xingyan [7] applied gesture recognition using fuzzy C-means algorithm, He has used image processing methods to transform the raw image into the feature vector. The feature vector is created by applying segmentation using HSV color model and then he has reduced the noise, the feature vector of the image is thirteen parameters long. The first feature is the aspect ratio of the hand’s bounding box as decided by the authors in [7]. The last 12 features are values representing coarse parameters of the image, where each grid cell is the mean gray level value in the 3 by 4 block division of image. Each of the 12 values calculated by the mean value of 3 by 4 partitions, which represent the mean of the brightness value. The classification phase is applied using a recognition algorithms based on the Fuzzy C-Means (FCM) algorithm, he had used samples taken from 6 persons with 6 gestures each and achives a time of 2-4 seconds with 86 % recognition rate.

3. OVERALL APPROACH

Our system is composed of four main segments in order to recognize the input gesture, these stages are summarized in Figure (3), and the details of these stages are as follows:

3.1 Collect the Input

In this phase, the input gestures are predisposed and the gesture poses are decided along with their meaning, and the database is creating which contains the different gestures poses with many samples per gesture, the number of the samples are limited by the speed and accuracy of the system, as the samples increased, the system speed decreased and the accuracy increased, and vice versa. We have chosen our database as in Figure (4) along with their meaning.

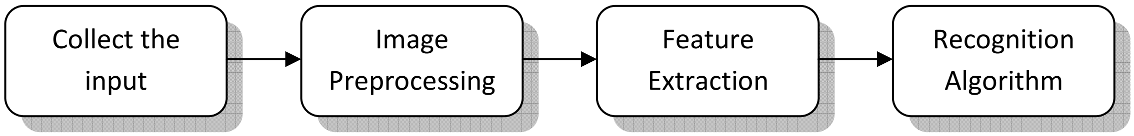


FIGURE 3: Overview of Gesture Recognition System.

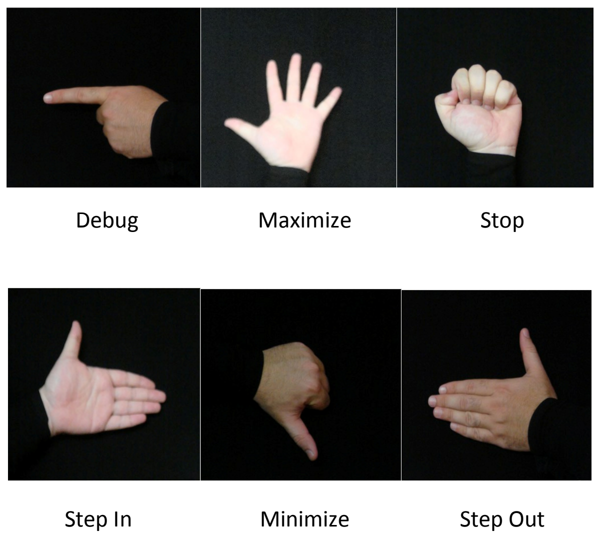


FIGURE 4: System's Database Vocabulary.

As seen by Figure (4), the database contains six gesture poses, which represent the target recognition decision which may belong to any of these six gestures if the new presented gesture is recognized, else, the system will announce that the ne presented gesture is not-known, these six gestures that implied in the database have many samples for each gesture which are six samples for each, as the number of samples increases the system accuracy increases and the testing time increases which effects badly on the overall speed and performance of the system.

3.2 Image Preprocessing
In this phase, we have predisposed the input gestures for both training and testing purposes; this phase is divided down into the following sub phases.

3.2.1 Segmentation
In this phase, we have applied the segmentation operation to segment the hand area in the input gesture and isolate it from the background, all of the gesture systems depend on the perfect segmentation of the hand gesture region, there are two main methods for segmentation: first method is by using HSV model; which deals with the color pigment of the human skin, the different ethnic groups have a significant property which is the different in the skin color which is represented by the pigment concentration difference which affect the saturation of the skin [7]. The color of the skin, on the other hand, is roughly invariant across ethnic groups [7]. By deciding a range for each of H, S and V parameters, a good segmentation method can be achieved; Xingyan Li [8] have decided certain values for H and S parameters only. The other method used for segmentation operation is by using clustering algorithms or thresholding technique, these algorithms are suitable mainly for homogeneous or uniform background that has no cluttered

objects and the hand is the prominent object, we have applied HSV mode of segmentation for splitting the hand area from the image gesture with some threshold value, after image segmentation; we have to normalize the segmented image gesture in order to obtain gestures database that are invariant against position, scale, and rotation, this will speed up the recognition operation and reduce the number of gesture samples stored in the database.

3.2.2 Edge Detection

The areas in images with high contrast that has an intensity jump between the connected pixels are called the edges, the non-edge areas do not has any intensity difference, noise also has some intensity difference, so, edge detection algorithms preserve the important information in the images that represents the edges and reduces the useless information which is classifies as non-edges pixels, there are two types of edge methods, the gradient (first derivative) and the laplacian (second derivative) as Figure (5), the gradient seeking for the maximum and minimum values and by using of thresholding techniques it classifies the edges and this methods is used when the intensity is high and is changed rapidly which provide a significant contrast, but when the gray level changes slightly from dark to light or vice versa the second derivative is used; the second derivative for maximum value is zero value which is the principle of the laplacian edge detector, this methods uses zero crossing to find the edges and it is suffers from false edges produced because of noise so it need special treatment like blurring the image before applying laplacian edge detector, we used this laplacian to find the edges.

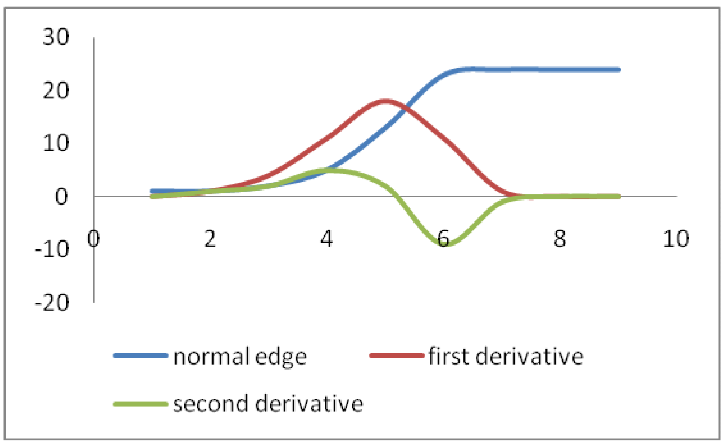


FIGURE 5: Edge Methods Categories.

As seen by Figure (50), the maximum value hereinabove is the edge in case of first derivative methods; this edge can be located by using of thresholding technique which produces different edge location depending on the threshold value and the edge thinness depends on the correct selection of the threshold as well, but in second derivative; the intersection with x-axis is the edge produces a unique and non-duplicated edge.

3.2.3 Normalization

In this phase of image preprocessing, the gesture is trimmed to get rid of the unnecessarily area that surrounding the gesture area, this is done by removing this useless area from four directions.

3.3 Feature Extraction

After the preparing of the image and segmenting of the hand gesture, a black-white image is created and represented the hand pose inset, the feature extraction phase will start, the overall feature vector size is 625 elements which are the brightness values of each block in the gesture, these features are stored in the database and the same algorithm is applied on the 36 gesture database, the actual feature vector size that stored in the database is more smaller than the calculated feature vector since the existence of zero values, this fact will be discussed hereafter.

3.4 Recognition Algorithm

When a new input presents to the system for classification, the input gesture is treated as prescribed sequence and extracting the features of this input gesture, the feature vector of the new presented input is compared against the stored database feature vectors using our new algorithm for matching purpose, each two brightness values are considered equal in case of both have no brightness value which means the block which is 5x5 (25 pixels) is black and this lead us to matching between two black areas, and in case of the existence of the brightness value, i.e. >0, we apply some threshold for considering the two blocks are equal, we use threshold of one to obtain maximum flexibility of the system in case of changing the inset of the same gesture, this matching algorithm yields good recognition rate comparing with other prescribed methods because the recognition rate depends on two decisions, the black matching and the brightness matching.

4. EXPERIMENTAL RESULTS

As we explained before, we trained the system with six different gestures each of six samples, these gesture images have undergone in a serial operations in order to extract the features.

First step is the segmentation operation, this operation is required for splitting the hand region, the segmentation is applied using HSV color space, the input gesture is RGB color space and converted to HSV color space using the following equations:
Let p(x, y) represents the input pixel with R, G, and B components, and let p'(x, y) with H, S, and B components, the following steps for converting p(x, y) to p'(x, y)[12]:

Step1: calculate M from, M = max(R, G, B)
Step2: calculate m from, m = min(R, G, B)
Step3: calculate r from, r = (M-R)/(M-m)
Step4: calculate r from, g = (M-G)/(M-m)
Step5: calculate r from, b = (M-B)/(M-m)
Step6: calculate V output from, V = max(R, G, B)
Step7: calculate S output from, if M = 0 then S = 0 and H = 180 degrees
if M <> 0 then S = (M - m) / M
Step8: calculate H output from, if R = M then H = 60(b-g)
if G = M then H = 60(2+r-b)
if B = M then H = 60(4+g-r)
if H >= 360 then H = H - 360
if H < 0 then H = H + 360
Step9: output H, S, and V color space, where H in the range [0,360], S and H in the range [0,100]

We decided a range of values for each of H, S and V so it accepts the pigment of human skin, let is the input gesture colored image location (x, y), and let H(x, y), S(x, y) and V(x ,y) are the H, S and V bands for the HSV color space for the input gesture image at location (x, y); let M be the output binary image, we set M(x, y) to 1 when Hmin<, H(x ,y)<Hmax and Smin<, S(x, y)<Smax and Vmin<, V(x ,y)<Vmax ; otherwise, we set M(x, y) to 0; Figure (6) is the output of applying this technique on Figure (4).

After this phase, hand boundary required to be last step in the image preprocessing, laplacian edge detector is applied as we explained before, the mask shown in Figure (7) is used and produces the output shown in Figure (8).

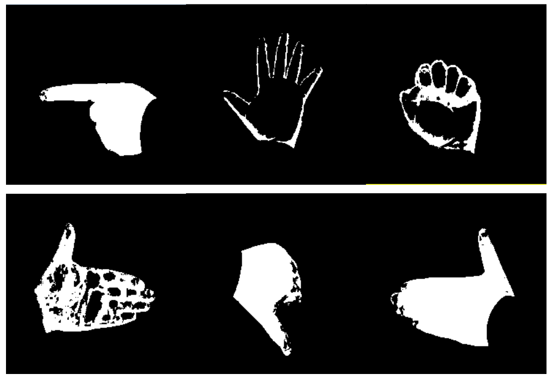


FIGURE 6: Binary Images.

0	1	0
1	- 4	1
0	1	0

FIGURE 7: Laplacian Mask.

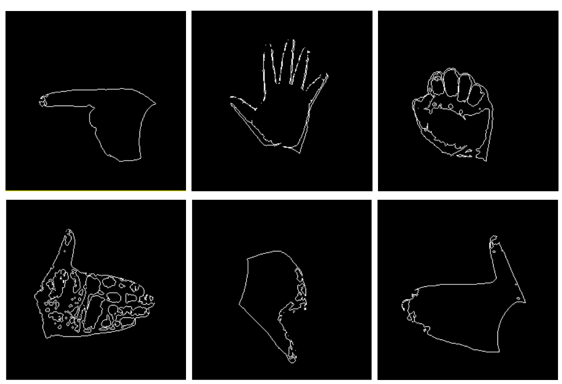


FIGURE 8: Edges of the Gestures.

Now, normalization operation is applied for removing the unwanted area of the gesture image, Figure (9) shows the application of normalization operation for one gesture of the samples gestures, the rest gestures are the same.

After this point, the gesture image is ready for feature extraction, as we said, the image output is 128x128 pixels, and block size is 5x5 pixels, so, each gesture image can produce 625 feature vector represents the features of this gesture, Figure (10) shown the values computed from the output gesture in Figure (9).

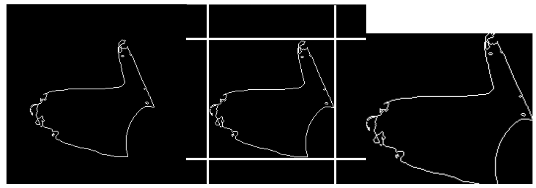


FIGURE 9: Normalization Operation.

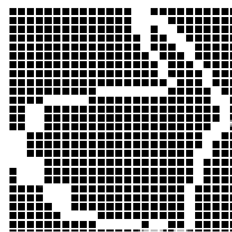


FIGURE 10: Features Calculation via Dividing the Gesture.

As seen in Figure(11), out of 625 features just 98 are white and stored and the other are neclected which represents 15.68 % stored features out of 625, and 84.32 % is neglected, Equation (1) shows the mathematical implementation for the gesture division.

$$B = \sum_{k1=-2}^{2} \sum_{k2=-2}^{2} M(k1+i, k2+j) \qquad (1)$$

Where i,j takes the values 2..height-3, 2.. width-3 respectively.
B: is the output brightness value.

After getting these features, and at the time of recognition during testing stage, the features of the testing gesture are calculated and compared with database features using our suggested algorithm, and the highest matching score is passed along with it is gesture and meaning, the algorithm is described below:

Consider D(i,k) is the database feature k of gesture i, and T(k) is the feature k of the testing image, M(k) is the matching status of feature k between both database gesture i and input testing gesture, for each gesture i, if D(i, k) is black area and T(k) is black area, the set M(k) as matched status, if the brightness value of D(i, k) > Threshold and T(k) > Threshold, then set M(k) as matched status, otherwise, set M(k) as non-matched status, after that the number of matched state is calculated and the recognition percentage with this database gesture is calculated via Equation (2).

$$Recognition\ Percentage = \frac{number\ of\ matched\ status}{total\ number\ of\ blocks} * 100\ \% \qquad (2)$$

And when the matching percentage exceed 85 %, the algorithm stops immediately for saving the time and matching is found, if not, the algorithm stays running until all gestures database are matched and the maximum matching percentage is passed.

5. TESTING AND VALIDATION

We test our system with 24 different samples, 4 gestures for each pose(total of 6 poses), and the recognition rate was 91 %, our system recognized the gestures with different transformations and achieve this percentage of recognition, the translation and scaling solved by normalization operation which produces a unified gesture image boundaries, rotation is managed here by brightness value and blocking method which allows some flexibility in gesture slope, the testing gesture is undergone in same serious of operation to get the features, after that these extracted features will be compared with the stored database features, the maximum matching score gesture is the equivalent gesture and it is meaning represents the meaning of the tested gesture, Figure (11) shows the testing gestures with their maximum matching probability.

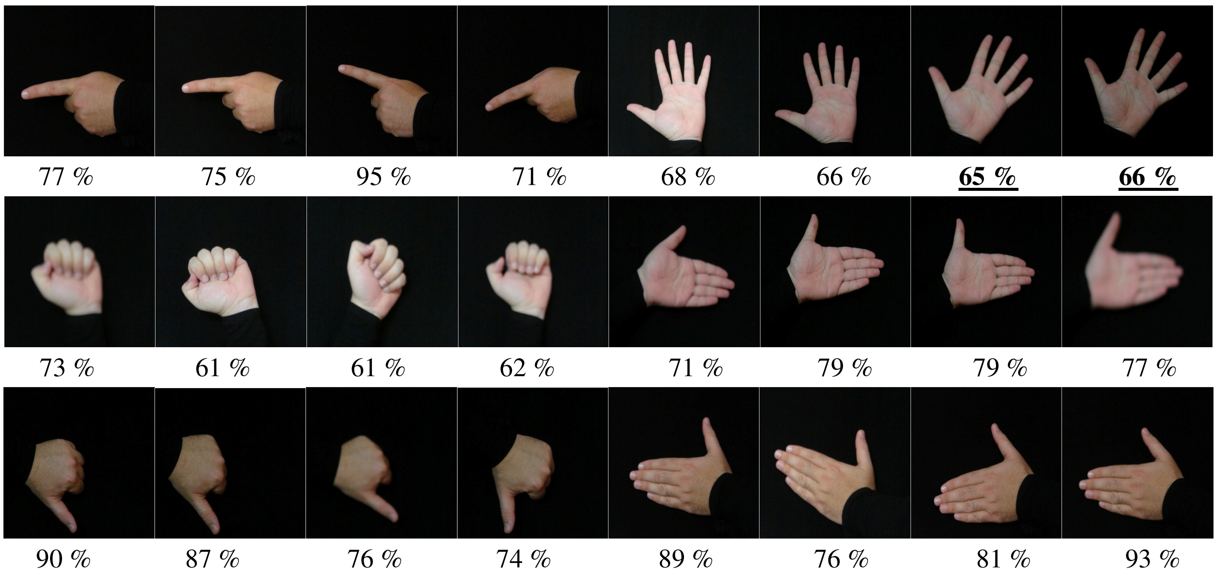


FIGURE 11: Recognition Percentage for each Tested Gesture.

In Figure (11), the underlined and bolded recognition percentage represents the non-recognized gestures by the system, they have highest probability but referring to wrong gesture. You can notice the rotation in these testing gestures and the system still recognizes these gestures. Figure (12) represents the matching chart for two selected testing gestures from Figure (11) which have matching percentage %93, %65 respectively, and first one is recognized gesture and the second one is non-recognized gesture.

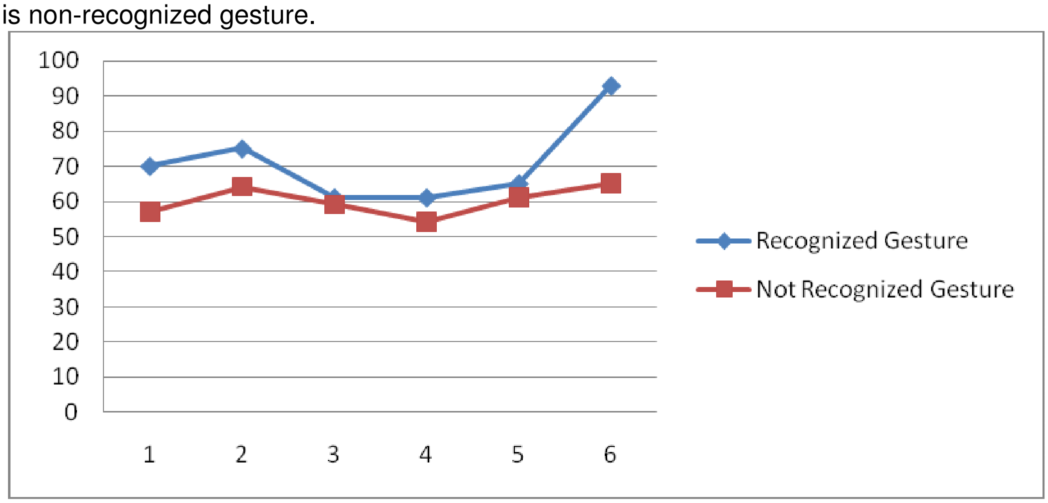


FIGURE 12: Matching Chart for Two Selected Testing Gestures.

In Figure (13), we have applied our recognition algorithm against all database feature vectors and the recognition rates are shown herein below, the aim is to reveal the prominent of the recognized gesture over all other gestures.

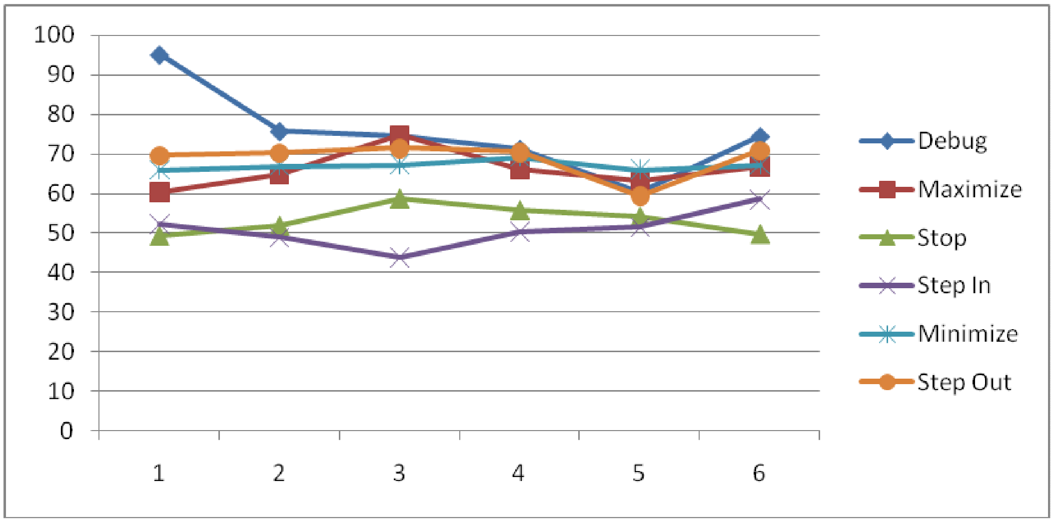


FIGURE 13: Matching Single Gesture against all Database Gestures.

6. CONSLUSION & FUTURE WORK

Hand gesture segmentation using HSV color model provide a promising results, this segmentation rely mainly on the pigment color of the human skin regardless the illumination which can be overcame by setting correct parameters for this model, also the seeking for skin color considered to be an efficient technique specially when the input image is cluttered with many objects, many researches commenced to enhance and develop the interaction between the human and human-made machines, template matching uses raw data and affected by rotation and transformation changes in addition to the illumination changes, in order to overcome this problem we need to increase the number of samples per gesture which indicates the increasing of the database size and in turn the processing time as well, the orientation histogram provides a promising solution for illumination changes due to the edge direction which is not affected with illumination variance, nevertheless, the hand object must dominate [6] the input gesture image and the database vocabulary must be chosen to avoid some confusion [6] in testing phase which may provide same testing result for different input gestures, by using of neural network the number of samples must be limited due to training time and the training data must be separable, in our method we have built a recognition system can handle a good degree of rotation via partitioning technique which partitions the hand gesture into blocks, and then we have applied the suggested testing algorithm.

Overall, feature selection is an important issue for gesture recognition and it is considered to be a crucial to the recognition algorithm, image preprocessing steps also important because the perfect steps can promise a unique and small feature vector which will reduce the number of samples in the database and speed up the recognition time.

In this study we have achieved 91 % recognition rate using different gestures at different rotation angles but using same conditions of illumination against uniform background, we have trained our system with 60% of gestures and tested our system with 40%, in the future work one can use the non-uniform background instead of uniform background.

7. REFERENCES

1. Wikipedia Internet Web Site

2. Internet Web Site. Available at:http://commons.wikimedia.org

3. S. Naidoo, C.W. Omlin, M. Glaser, *"Vision-Based Static Hand Gesture Recognition using Support Vector Machines"*. Department of Computer Science, University of the Western Cape, South Africa, 1999

4. K. Symeonidis. *"Hand Gesture Recognition using Neural Networks"*, Master Thesis, School Of Electronic And Electrical Engineering, 2000

5. R. Brunelli, T. Poggio. *"Face Recognition: Features versus Templates"*. IEEE Transactions on Pattern Analysis And Machine Intelligence, 15(10):1042-1052, 1993

6. W. T. Freeman, M. Roth. *"Orientation Histograms for Hand Gesture Recognition"*. Mitsubishi Electric Research Laboratories, Cambridge, Ma 02139 USA, 1994

7. X. Li. *"Gesture Recognition based on Fuzzy C-Means Clustering Algorithm"*. Department Of Computer Science The University Of Tennessee Knoxville, 2003

8. X. Li. *"Vision Based Gesture Recognition System with High Accuracy"*. Department of Computer Science, The University of Tennessee, Knoxville, TN 37996-3450, 2005

9. J. Wachs, U. Kartoun, H. Stern, Y. Edan. *"Real-Time Hand Gesture Telerobotic System using Fuzzy C-Means Clustering"*. Department of Industrial Engineering and Management, Ben-Gurion University of the Negov, 1999

10. J. Triesch, C. Malsburg. *"Robust Classification of Hand Postures Against Complex Backgrounds"*. IEEE Computer Society, In Proceedings of Second International Conference On Automatic Face And Gesture Recognition, 1996

11. T. Yang, Y. Xu. *"Hidden Markov Model for Gesture Recognition"*. The Robotics Institute Carnegie Mellon University Pittsburgh, Pennsylvania 15213, 1994

12. J. J. Phu, Y. H. Tay. *"Computer Vision Based Hand Gesture Recognition using Artificial Neural Network"*. Faculty of Information and Communication Technology, University Tunku Abdul Rahman (Utar), Malaysia, 2006

13. H. B. Amor, S. Ikemoto, T. Minato, H. Ishiguro. *"Learning Android Control using Growing Neural Networks"*. Department Of Adaptive Machine Systems Osaka University, Osaka, Japan, 2003

14. M. Swain and D. Ballard. *"Indexing via Color Histograms"*. In Proceedings of Third International Conference on Computer Vision, 390-393, 1990

15. S. Venkataraman, V. Gunaseelan. *"Hidden Markov Models in Computational Biology"*.lectures in HMM

16. The AF Research Laboratory. *"Language and Cognition"*. Elsevier, Neural Networks, 22: 247-257, 2009

17. H. Gunes, M. Piccardi, T. Jan."*Face and Body Gesture Recognition for a Vision-Based Multimodal Analyzer"*. Computer Vision Research Group, University of Technology, Sydney (UTS), 2007

18. Y. Lu, S. Lu, F. Fotouhi, Y. Deng, Susan J. Brown. *"A Fast Genetic K-Means Clustering Algorithm"*. Wayne State University, Kansas State University Manhattan, USA, 2000

19. B. Heisele, P. Ho, T. Poggio. *"Face Recognition with Support Vector Machines: Global versus Component-based Approach"*. Massachusetts Institute of Technology Center for Biological and Computational Learning Cambridge, 2001

20. K. Jain, R. P.W. Duin J. Mao. *"Statistical Pattern Recognition: A Review"*. IEEE Transactions on Patterns Analysis and Machine Intelligence, 22(1):4-35, 2000

21. S. S. Keerthi, O. Chapelle, D. DeCoste. *"Building Support Vector Machines with Reduced Classifier"*. Complexity, Journal of Machine Learning Research, 8:1-22, 2006

22. Abin – Roozgard. *"Convolutional Neural Networks"*. lectures in Neural Networks

23. Y. P. Lew, A. R. Ramli, S. Y. Koay, A. ali, V. Prakash. *"A Hand Segmentation Scheme using Clustering Technique in Homogeneous Background"*. Student Conference on Research and Development Proceedings, Shad Alam, Malaysia, 2002

24. C.C. Lo, S. J. Wang. *"Video Segmentation using a Histogram-Based Fuzzy C-Means Clustering Algorithm"*. Institute of Information Management, National Chiao-Tung University, Computer Standards & Interfaces, 23:429–438, 2001

25. S. Marcel, O. Bernier, J. Viallet, D. Collobert. *"Hand Gesture Recognition using Input–Output Hidden Markov Models"*. France Telecom Cnet 2 Avenue Pierre Marzin 22307 Lannion, France, 1999

26. C. Karlof, D. Wagner. *"Hidden Markov Model Cryptanalysis"*. Computer Science Division (EECS) Univertsity of California Berkeley, California 94720, 2004