BRIGHTNESS FACTOR MATCHING FOR GESTURE RECOGNITION SYSTEM USING SCALED NORMALIZATION

Mokhtar M. Hasan[1] and Pramoud K. Misra[2]

[1]Department of Computer Science, Banaras Hindu University, Varanasi, India
mmwaeli@gmail.com

[2] Department of Computer Science, Banaras Hindu University, Varanasi, India
pkmisra@gmail.com

ABSTRACT

The rich information found in the human gestures makes it possible to be used for another language which is called the sign language, this kind of intuitive interface can be used with human-made machines/devices as well, we herein going to introduce a new gesture recognition system based on image blocking and the gestures are recognized using our suggested brightness factor matching algorithm, we have applied two different feature extraction techniques, the first one based on features extracted from edge information and the other one based on a new technique for centre of mass normalization based on block scaling instead of coordinates shifting; we have achieved 83.3% recognition accuracy in first technique with significant and satisfactory recognition time of 1.5 seconds per gesture, and 96.6 % recognition accuracy with recognition time less than one second by eliminating the use of edge detector which consumes time, this paper focuses on appearance based gestures.

KEYWORDS

Brightness Recognition, Gesture Recognition, Template Matching, Image Segmentation, Edge Detection, Laplacian Edges, Gray Conversion, Normalization, Centre of Mass, Feature Extraction

1. INTRODUCTION

The ubiquitous computers are used in our daily life and became the inevitable mate, the encumbered input devices such as keyboard and mouse and even the touch screen are limited the human capabilities for the communicating with these devices, new interface is demanding for the human-made machines, therefore, the vision based input devices attracted the attention too be the replacement of such former boring devices, Figure 1 shows the usage of gestures for warplane directing.

Figure 1. Signaller directs the warplane parking [1].

The gestures can originate from any bodily motion or action like hand, head, figures and any bodily organ that can express an act. Gestures also may represent the compressed information [2] that can be decoded between the sender and receiver in order to adopt the gesture meaning, we can interpret the hand fist as any predefined meaning between the involved objects, this agreement can be translated to take place between the human and the human-made machines or even devices.

Many challenges facing this gesture system, we can classify the challenges into two major types, these types are listed below:
(1) Posture Challenges
(2) System Challenges
Hasan and Misra [3] address three types of challenges, illumination challenge that affects the extraction of the correct skin-like region, transformation challenge which means that the hand region may fall within any portion of the input image, and background challenge which mix many objects beside the hand object in the gesture image, these challenges may classified as posture challenges, in the other hand, the classification speed, accuracy and human-skin colour detection are addressed by Hasan and Misra [4], speed is very important for real time gesture system, no noticeable time delay [4] should be between the commence of the gesture and the trigger of the command and should be with acceptable accuracy, and also the system should has the ability to operate on different ethnic group [4], those kind of challenges may classified as system challenges.

In this paper we have applied an appearance based model for hand detection and recognition, we have used the gray scale color model that converted from the input coloured images, and we have applied our designed recognition algorithm based on local computing of brightness value.

2. RELATED WORK

Sebastien Marcel [5] has opened a single block called "active window" for user commands, he applied hand posture detection and recognition using the LISTEN system and applied using a neural network called constrained generative model (CGM), he mapped a body-face space as detected by LISTEN system, he opened an active window for detecting the user intension for command issuing, When a skin color goes inside this "active window", the recognition algorithm is triggered, he used 6 different posture classes with many sample and applied for

both uniform and cluttered background, he achieved 93.6 % with uniform background and 75.45 % with cluttered background by using his own database.

Joshua R. New [6] built a recognition system for hand movement tracking for four selected movements, which are circle, square, left/right, and up/down, by sitting in front of a computer table with an off-the-shelf camera is attached pointed downward to capture the underneath area where the human hand suppose to lay, and by using the center of mass technique to locate the hand in the captured image, 12 frames represent the full gesture meaning out of the four gestures as mentioned hereinabove, and a centeriod value has been calculated for each frame, those 12 centeriods represent a trace for what the user did during last 3 seconds, and then a 24 feature vector size is created from those 12 frames to be classified by Using Simplified Fuzzy Adaptive Resonance Theory Map, as he mentioned; the main limitation is as soon the 3 seconds start to capture a hand gesturing, the user can not quit or change the gesture or else, misclassification by the system is triggered.

The authors in [7] have applied a gesture recognition system by using centeriod profile, they divided the difference images into blocks of 8x8, and they calculates the entropy for each sub bock, the difference images with high entropy indicate the existence of a gesture, after that the hand region is detected by the maximum entropy since entropy can expresses color information, the hand contour is extracted by HSI color model and the recognizing of gesture is applied by the centeriod profile which the normalized distance from the centeriod is used to expresses the hand shape, they achieved 95 % recognition rate per person.

3. SYSTEM ARCHITECTURE

The architecture of gesture recognition system is divided into three main phases, the image pre-processing, feature extraction, and recognition algorithm, the first phase emphasis on removing all unnecessary information from the input posture like noise information, background information and other non-related objects like human face and his second hand in case of a single hand resonation system, the vital step in image pre-processing is the segmentation operation which plays a major role for showing the hand posture and removes all other non-related objects, we can apply edge detection after this step in order to get the boundary of the extracted hand but this causes to loss much information since the inner and outer areas of the hand segment will be treated the same while but in the other hand the processing time is reduced.

The feature extraction phase responsible on converting the vision version of the extracted hand into a numerical version, this latter format enable the compute/robot to process and understand the hand shape and this is the way that he understands his surrounding environment.

Recognition algorithm is used to establish a technique for finding the correct gesture among the stored database gestures and it is used at the time of testing, i.e. it seeks for a corresponding stored image gesture for the numerical features; Figure 2 shows the gesture system.
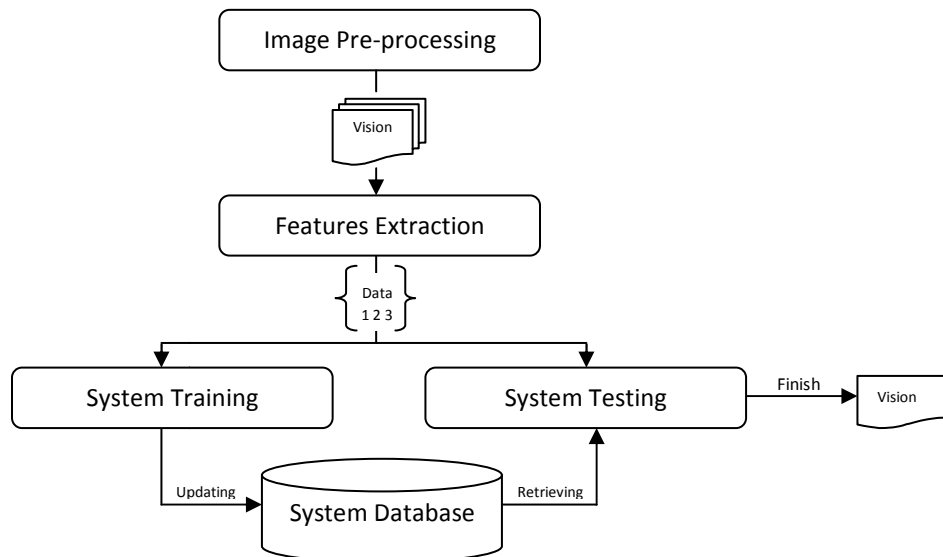
Figure 2. Gesture Recognition System Paradigm.

4. SYSTEM IMPLEMENTATION

We have applied the system using six different postures; 5 samples per posture are used to build the database and same number for testing, we have made this comparative study to present out new normalization technique and reveal its efficiency.

4.1. Block Scaling Normalization

The present techniques for normalization using centre mass is by shifting the coordinates so that the centeriod of the object will match the point of origin of XY axis, this technique produces the unbalanced portions of the segmented hand around the axis which depends on the existence of subject mass and hence, we need to rescale it to obtain a matching corresponding between the new presented object and the stored objects since the database and the new presented posture can be originated from any hand movement, Figure 3 illustrates ocularly this problem.
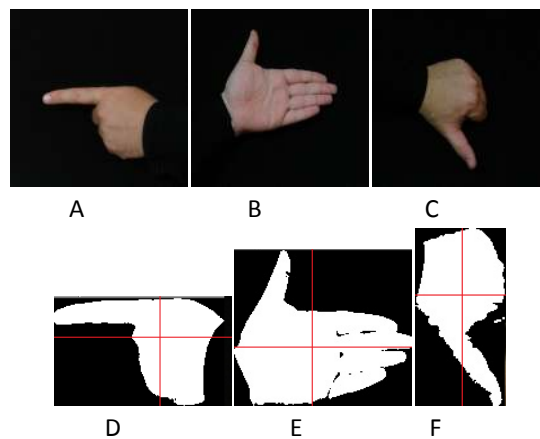
A, B, and C are the original postures, D, E, and F are the corresponding centre mass.

Figure 3. Classical Normalization.

As seen in Figure 2, the normalization does not preserve the object dimensions and shift the coordinates of the object, in latter figure; suppose we want to match the upper left block of image D with same block of E, this cannot be done unless rescaling has to be done first, and the same for matching all object with other objects.

Our suggested block scaling normalization preserve the dimensions and the time as well, one scaling is enough to maintain a unified posture dimensions for whole the database, Figure 4 illustrates the new presented normalization applied on D, F, and F of figure 3 respectively.

Figure 4. Block Scaling Normalization.

As seen in latter figure, all postures have fixed to a uniform dimensions and so the matching will be more quick and does need any rescale, now we can move forward to our proposed algorithm.

4.2. Image Pre-processing

The input image is undergone in a serial of processing steps, these steps are essential to remove the non-relevant hand objects and maintain the hand object, the selection of efficient segmentation algorithm plays a major role for extraction of this hand object, we have employed a uniform background because we want to examine the performance of our suggested method regardless of the existence of a background, and we want to check the system validity after the extraction of a hand segment, we have applied a threshold segmentation using gray level colour model.

Noise removing techniques can be employed in order to remove the noises which may happened due to light illumination affection or overlapping of some objects that produces a skin-like colour.

5. POSTURE FEATURES

We have applied our system using two different features as listed below, the database are formed by a feature vector for each single posture sample, this feature vector is calculated using our suggested method of brightness features which based on calculation of the local brightness of the local blocks of the hand segment; each block is 5x5 pixels and hence, the image produces 625 feature vector size, for more storage efficient and high processing speed; we can store the brightness values of these features and neglecting the non-bright features, the following algorithm explains these steps of brightness calculation.

Algorithm1, Brightness Value Calculation
Input: Hand Segment Image I (h, w), b is block size
Output: Feature Vector F (n)

Method:

Step 1: [initialization]
r = 0, c = 0, w1 = 0, h1 = 0, NEXT_FEATURE=0
n = (w/b) * (h/b)

Step 2: [feature iteration]
Repeat to step [5] while NEXT_FEATURE < n
Sum=0

Step 3: [block calculation]
Repeat to step [4] while r < b

Step 4: [summation]
If I(r+h1, c+w1) is bright then increment sum by 1
Increment c by one,
If c >= b then set c=0, Increment r by one
Return to step 3

Step 5: [update feature vector]
If sum > brightness threshold then F (NEXT_FEATURE) = sum else F (NEXT_FEATURE) = 0
Increment NEXT_FEATURE by one
Increment w1 by b
If w1-b >= w then set w1=0, increment h1 by b
Return to step 2

Step 6: [output]
Output the feature vector F with NEXT_FEATURE features available.
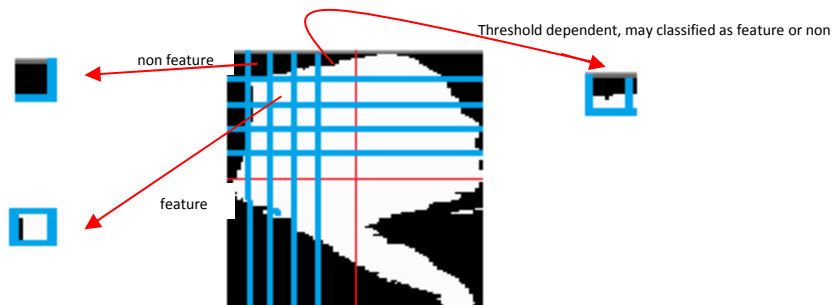End

Figure 5 shows the selection of the features.

Figure 5. Feature Selection.

5.1. Normal Features

The calculation of these features are based on the edges of the segmented hand, so that the hand boundaries will be included in the calculation and the inward and outward hand are considered to be same since there is no brightness in these areas.

5.2. Normalized Features

The edge detection method is not applied here after the segmentation operation, so the inward and outward portions of the hand object are considered to be different and this is an additional advantage.

6. TESTING AND VALIDATION

In order to show the strength of our suggested method, we have applied this algorithm by using 5 samples per pose out of 6 poses which mean 30 samples, some samples has some degree of rotation; some of these samples are misclassified by the first techniques but recognized correctly by the second method, Figure 5 shows the database for two selected poses.
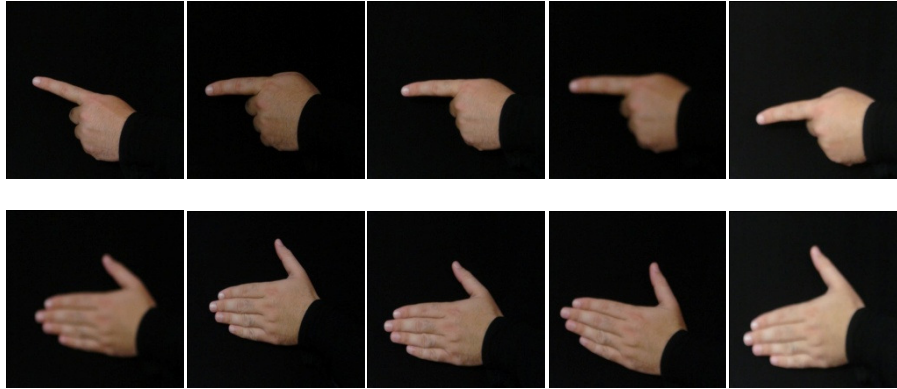

Figure 5. Training database for the selected postures.

Figure 6 shows the two selected poses for testing purpose which include some degree of rotation.
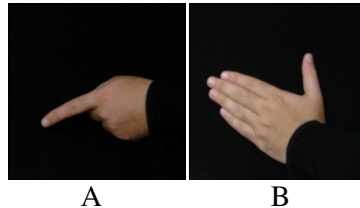

A   B

Figure 6. Two selected postures for testing purposes.

If we make a ocularly glance; we can notice the degree of rotation of these two selected poses and the system did not trained for such kind of rotation, these two selected testing poses are misclassified by using first technique as we mentioned with standard derivation of 2.949576 and 2.65832 respectively which means the localization and overlapping of recognition accuracy within a small area which makes it difficult to recognize, while the standard derivation by using second method is 6.055301 and 5.00999 respectively which gives a wider area for distributing of recognition accuracy, Figure 7 and Figure 8 show the recognition accuracy after applying each of normal and block scaling normalization methods.
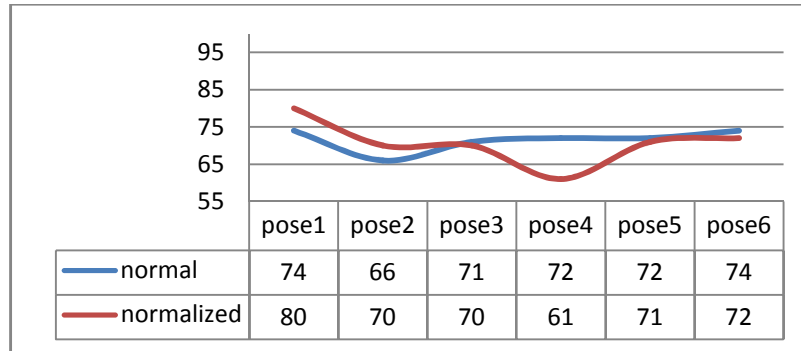
	pose1	pose2	pose3	pose4	pose5	pose6
normal	74	66	71	72	72	74
normalized	80	70	70	61	71	72

Figure 7. Recognition Accuracy (%) for Image A in Figure 6.

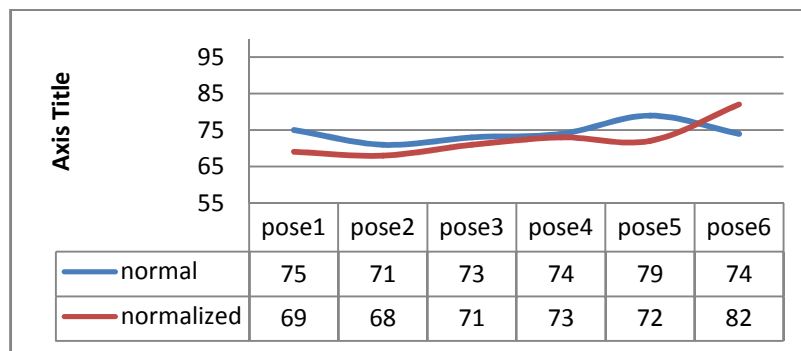|	pose1	pose2	pose3	pose4	pose5	pose6
normal	75	71	73	74	79	74
normalized	69	68	71	73	72	82

Figure 8. Recognition Accuracy (%) for Image B in Figure 6.

As noticed in Figure 7, in case of normal features (first technique) there is no distinguish peak point, in the contrary the normalized features can be distinguished by finding the peak point and the corresponding posture from the database is announced, Figure 8 the same.

We also traced single gesture and found that the recognition accuracy using second method is better than first method; this efficiency comes in term of high and noticeable peak and the big distance from other non-recognized postures as seen in Figure 9 below.
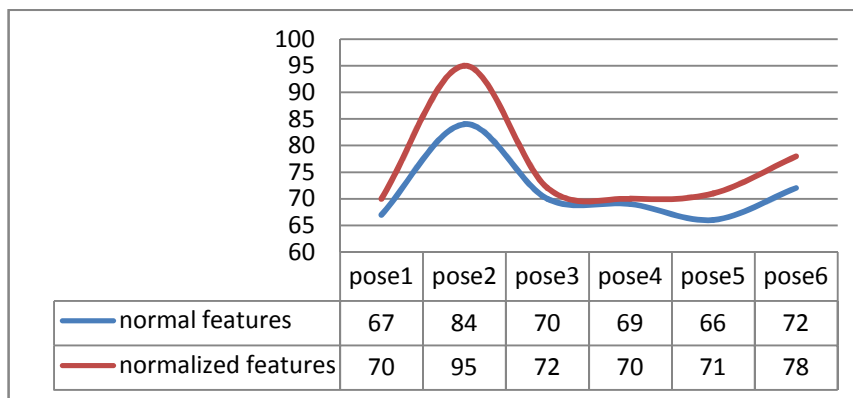
	pose1	pose2	pose3	pose4	pose5	pose6
normal features	67	84	70	69	66	72
normalized features	70	95	72	70	71	78

Figure 9. The dominant of recognized gesture peak point over non-recognized (%).

We also have calculated the average of five maximum values per posture; i.e. we have five testing samples per posture type and those five samples are recognized and then averaged, this is repeated for the whole six postures regardless if it is recognized or not in order to explain the altitude of peak point, this is done in Figure 10.

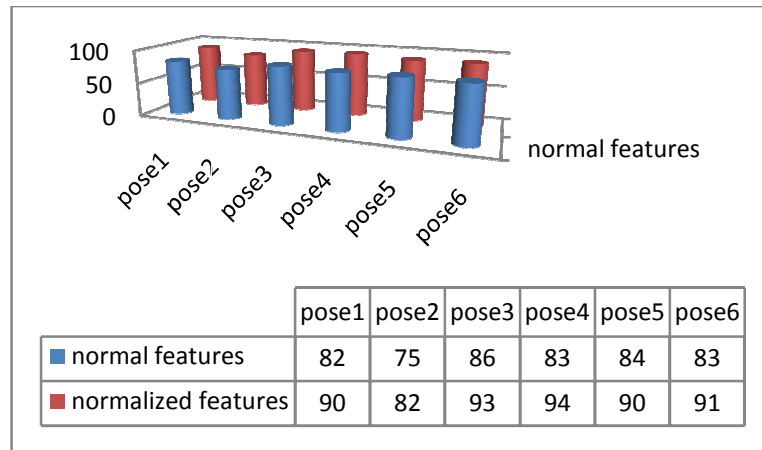|	pose1	pose2	pose3	pose4	pose5	pose6
■ normal features	82	75	86	83	84	83
■ normalized features	90	82	93	94	90	91

Figure 10. The averages of each recognized posture type separately.

As noticed, the second technique has an edge over the first technique which means the recognized pose can be distinguished more easily with a high peak point.

7. CONCLUSIONS

The vision based hand gesture embody a rich information that needed to be used in a correct way, we have presented two techniques for gesture system, first one based on edge information and other one based on segment information, the latter one shows a good experimental results after combine it with block scaling normalization because it uses as much data from the hand vision information.

The new implemented block scaling normalization produces a unified database features regardless of the object pose which may be a hand fist pose or hand outstretch pose or any other pose that may produce different dimensions, and by generalization the dimensions we have prepared the hand segment for the latter recognition algorithm.

We have applied our system with 5x5 block size, we can use other block size which may give better results, at least we have established the basic stone, the traditional pixel matching techniques have some lack since these techniques look for the corresponding pixel in one way or another, here, the 5x5 block size gives 25 pixels and the matching means looking for a pixel within 25 pixels which is more flexible with preserve the general shape of object since the matching decision is based on the whole blocks of the object and not just this block.

REFERENCES

[1]   Found at: article.wn.com

[2]   Sushmita Mitra, and Tinku Acharya , (2007), "Gesture Recognition: A Survey", *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications And Reviews*, Vol. 37, No. 3, pp311-324.

 [3]   Mokhtar M. Hasan, Pramod K. Misra, (2010) "Robust Gesture Recognition using Euclidian Distance", *IEEE International Conference on Computer and Computational Intelligence,* China, Vol. 3,pp.38-46.

[4]   Mokhtar M. Hasan, Pramod K. Misra, (2010), "HSV Brightness Factor Matching for Gesture Recognition System", *International Journal of Image Processing,* Malaysia, Vol. 4, No. 5, pp. 456-467.

[5]   Sebastien Marcel, (1999), "Hand Posture Recognition in a Body-Face Centered Space", *In Proceedings of Conference on Extended Abstracts on Human Factors in Computing Systems (CHI '99),* New York, USA.

[6]   Joshua R. New, (2004),   "A Method for Temporal Hand Gesture Recognition", *Knowledge Systems Laboratory,* Jacksonville State University, Jacksonville.

[7]   Jae-Ho Shin, Jong-Shill Lee, Se-Kee Kil, Dong-Fan Shen, Je-Goon Ryu, Eung-Hyuk Lee, Hong-Ki Min, Seung-Hong Hong, (2006),   "Hand Region Extraction and Gesture Recognition using Entropy Analysis", *International Journal of Computer Science and Network Security,* Vol.6, No. 2A,pp.216-222, February.

 [8]   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.

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

[10]   R. Brunelli, T. Poggio, "Face Recognition: Features versus Templates", IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 15, No. 10, October 1993.

[11]   William T. Freeman, M. Roth, "Orientation Histograms for Hand Gesture Recognition", *Mitsubishi Electric Research Laboratories,* Cambridge, Ma 02139 USA, 1994.

[12]   X. Li, "Gesture Recognition based on Fuzzy C-Means Clustering Algorithm", D*epartment Of Computer Science,* The University of Tennessee Knoxville, 2003.

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

[14]   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.

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

[16]   T. Yang, Y. Xu , "Hidden Markov Model For Gesture Recognition", *The Robotics Institute Carnegie Mellon University Pittsburgh,* Pennsylvania 15213, 1994.

[17]   Jiong J. Phu, Yong 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.

[18]   Heni 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.

[19]   M.Swain and D.Ballard, Indexing via Color Histograms, *International Journal of Computer Vision,* vol. 7, pp. 11-332, 1991.

[20]   S. Venkataraman, V. Gunaseelan, "Hidden Markov Models in Computational Biology", lectures in HMM.

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

[22]   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.

[23]   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.

[24]   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.

[25]   Anil K. Jain, Robert P.W. Duin, and Jianchang Mao, "Statistical Pattern Recognition: A Review", *IEEE Transactions on Patterns Analysis and Machine Intelligence,* 22(1):4-35, January 2000.

[26]   S. Sathiya Keerthi, O. Chapelle, D. DeCoste, "Building Support Vector Machines with Reduced Classifier, Complexity", *Journal of Machine Learning Research* 8 (1-22), 2006.

[27]   Abin – Roozgard, "Convolutional Neural Networks", lectures in Neural Networks.

[28]   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.

[29]   Chi-Chun Lo, Shuenn-Jyi 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.

[30]   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.

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