International Journal of Engineering Research and Development (IJERD) ISSN: 2278-067X
Recent trends in Electrical and Electronics & Communication Engineering (Page 17-25)
(RTEECE 08th – 09th April 2016)

# Appearance and Statistical Features Blended Approach for Fast Recognition of ASL

Bhavin Kakani[1], Hardik Joshi[2], Rohit Yadav[3], C Archana[4]

***Abstract:-*** *Hand gesture recognition has wide range of applications in many areas. The development of Gesture recognition is the witness of the human being effort aiming to communicate with electronic gadgets. This paper deals with an efficient algorithm to recognize the hand gestures representing the American Sign Language (ASL). All the features used are basically shape based assuming that the shape of human hand is same for all human being. The main features used are of two types: appearance based features and the correlation based features. Template matching uses the cross-correlation to recognize the similar pattern. The timing analysis of the proposed method is done for further optimization of the algorithm. The project work is divided into two parts: first is the gesture detection and second is the gesture recognition based on feature extraction.*

***Keywords:-*** *American Sign Language (ASL), Cross correlation, Euclidian distance, Histogram of Oriented Gradients (HOG)*

## I. INTRODUCTION

Now a days computers are playing an essential role in human life. Hand gesture recognition is an important area of computer vision and pattern recognition field as the most flexible part of human body, hands play an important role in human's daily life communication.

The development of Gesture recognition is the witness of the human being effort aiming to communicate with electronic gadgets. From the wide range of applications, this paper focuses on the hand gesture recognition for American Sign Language (ASL) symbols. American Sign Language is the dominant sign language of the Deaf community in the United States, in the English-speaking parts of Canada, and in parts of Mexico. So, we have used ASL images as input database to design the algorithm[1].

Numerous techniques are used for the hand gesture recognition. Based on the extra devices required, we can classify it into data glow based technique, color glow based technique and the vision based technique. Vision based hand gesture interface has been attracting more attentions due to no extra hardware requirement except Camera, which is very suitable for ubiquitous computing and emerging applications. Methods for vision based hand gesture recognition fall into two categories: 3D model based methods and appearance model based methods. 3D model may exactly describe hand movement and its shape, but most of them are computational expensive to use.

Recently there are some methods to obtain 3D model with 2D appearance model such as ISOSOM and PCA-ICA in [7][8].

In this paper, we focus on appearance model based method. There have been a number of research efforts on appearance based method in recent years. Freeman and Weissman recognized gestures for television control using normalized correlation [9]. This technique is efficient but may be sensitive to different users, deformations of the pose and changes in scale, and background. Cui and Weng proposed a hand tracking and sign recognition method using appearance based method [10]. Although its accuracy was satisfactory, the performance was far from real-time. Just et al introduce modified census transform into hand gesture classification [11]. For the purpose of classifying each gesture respectively, their method obtains fairly good results. While the performance in recognition experiments was not so satisfactory and the recognition result of different gesture great disparity. Elastic graphs were applied to represent hands in different hand gestures in Triesch's work with local jets of Gabor filters [12]. It locates hands without separate segmentation mechanism and the classifier is learned from a small set of image samples, so the generalization is very limited. The performance of vision based gesture interaction is prone to be influenced by illumination changes, complicated backgrounds, camera movement and specific user variance. Many researchers have made effective efforts to deal with these problems. Lars and Lindberg used scale-space color features to recognize hand gestures [13]. In their method, gesture recognition method is based on feature detection and user independent while the authors showed real-time application only under uniform backgrounds.

Hand gesture recognition system includes two phases: object detection and object recognition. Object detection is to try to find the position of a certain object from the entire image. Object recognition is to try to recognize a certain pattern that differentiates the object from the other objects.

In fig. 1, the general flow of algorithm is shown.

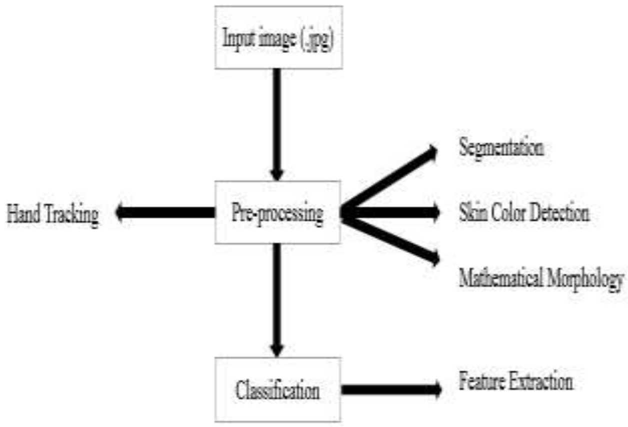


**Fig.1:** General Block diagram of the proposed algorithm

## II. DESIGN AND IMPLEMENTATION

### A. Creating database

A set of 140 images (ASL symbols) were initially clicked and stored in the memory as database images. Camera resolution was set at 0.3 Mega pixel and image dimensions are 640x480.Background is keep black so as to avoid and minimize the effect of shadow of the hand although some illumination variation is there in different sign images. J and Z are not considered as they are not static symbols in ASL.

Three different databases of 34 images were created (A-Y and 0-9) for experimental training purpose and some dummy images (other than database images) were taken for testing purpose.

### B. Object Detection

Object detection is the first and most important step in the image processing field. Pre-processing steps are included in the object detection phase. In pre-processing, basically captured image is modified in a way that it can be easily used in the further stages of the system. Hand detection in our method is an initial step of interaction. It's important for a gesture interface as it functions as a switch to turn on the interface. A number of processes were included in pre-processing steps [1].

### C. RGB to YCbCr Image Conversion

Since RGB images are prone to illumination variation so the initial step in proposed method is to change the color space from RGB to YCbCr. This color domain also helps in fighting against shadow effects. Cb and Cr components are used for skin color detection.[1][2].

In Fig. 2, the original database image and its processed YCbCr image is shown.

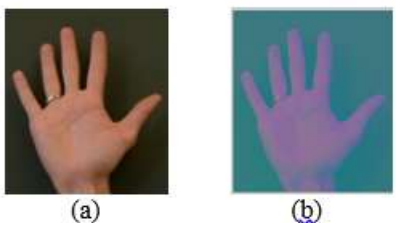


**Fig.2:** (a) Original image, (b) YCbCr converted image

For skin detection purpose minimum and maximum skin probabilities of processed image are calculated. Data using normalized skin probability used for producing gray threshold at zero which gives binary image.Fig. 3 shows grayscale and binary image.

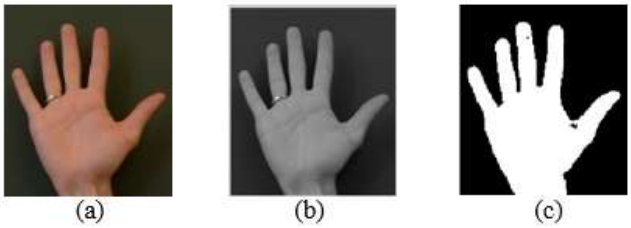


**Fig.3:** (a) Original image, (b) Gray scale image, (c) Binary image

### D. Noise in Binary Image

Noise in binary image can be of two types: 1. Noise in background region and 2. Noise in hand region. This noise can significantly affect the performance of the algorithm. Mathematical morphology was used to remove such noise in the binary image.

Morphological operation imclose applied on filtered image which performs closing the binary image and close operation is a dilation followed by erosion, using structuring element. Dilation is the process where the value of the output pixel is the maximum value of all the pixels in the input pixel's neighborhood. So, it will assign ‘1’ to the pixel if the neighbors are ‘1’. So, this operation was used to remove the noise from the hand region. On the other hand, Erosion is the process where the value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. So, this operation was used to remove the noise from the background.[3]

In fig.4 the effect of both these operations is shown


**Fig.4:** (a) Effect of Dilation, (b) Effect of Erosion

8-Component connectivity of binary image is calculated using bwlebel and regionprops is applied to find bounding box for white pixels of hand image and imcrop is applied to extract region of interest (ROI). Next step is to extract the features of each image to recognize at later part.

## III. FEATURE EXTRACTION AND OBJECT RECOGNITION

In the object recognition phase, various operations to be performed on the binary image. This entire process is also known as feature Extraction. Feature extraction process can be divided into two categories according to the features of image used. Fig.5 shows this classification.

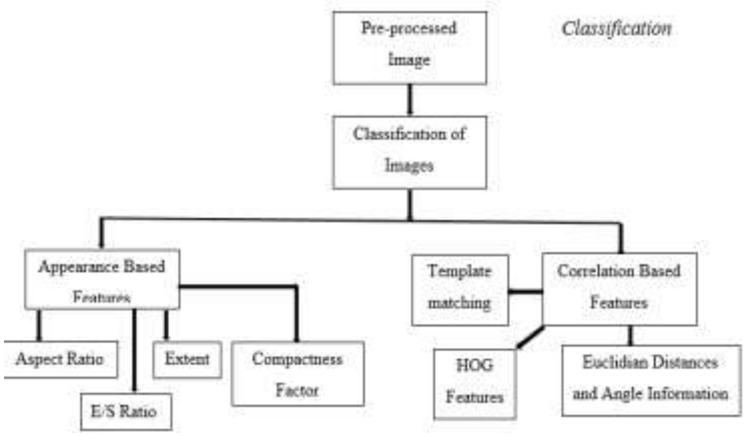


**Fig.5:** Processed image and bounding box around the hand region

Appreence based features involves Aspect Ratio, compactness factor and extent whereas cross correlation based features involves template matching, HOG features[5] and angleinformation of peaks. As in the later type, we need to compare test image with all the images,The appearance based features are performed for classification of symbols.

A. **Aspect Ratio(Orientation of image)**

A bounding box is made around the processed image and its width to height ratio (W/L) is calculated. If the ratio is less than 1 then it can be categorized as "vertical symbols" else "horizontal symbols" as mentioned in Fig.6. Also symbols with W/L<0.7 are categorized as "open symbol" and W/L>0.7 are considered "closed symbol"[2].

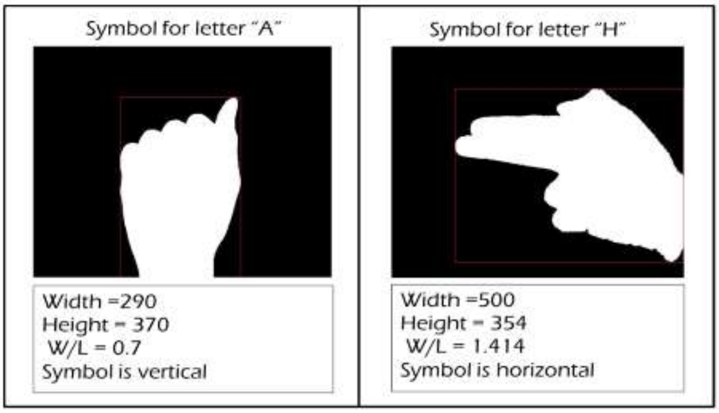


**Fig.6:** Processed image and bounding box around the hand region

After determining orientation of all the images, The entire database can be seperated into two groups and the remaining operations to classify the different gestures are performed on the group of vertical and horizontal images.

B. **Extent**

It is another feature based on the information obtained above. Extent is a ratio of the area of the hand region to the area of entire bounding box.

The reason to select this feature is that it is easy to apply and highly reliable and robust to classify particular gesture. Fig.7 shows the importance of this feature.

**Fig.7:** Bounding box across two different symbols

First symbol contains space between fingers, so its bounding box area is more as compared to the area of hand region, whereas second symbol doen't contain any space, so its extent value against bounding area would be almost unity.

C. **Compactness Factor**

The compactness measure of a shape, sometimes called the shape factor, is a numerical quantity representing the degree to which a shape is compact. It is determined by squared perimeter of the shape to its area. This also means that two patterns with almost the same squared perimeterto area ratio would exhibit the same compactness value. This value is not affected by the orientation and scaling the image which makes it a robust feature.[2]

$$compactness = \frac{perimeter^2}{4\pi \times \mathbf{area}}$$

This is the end for appearance based features and the classification of the gestures based on these features are shown in Fig.8

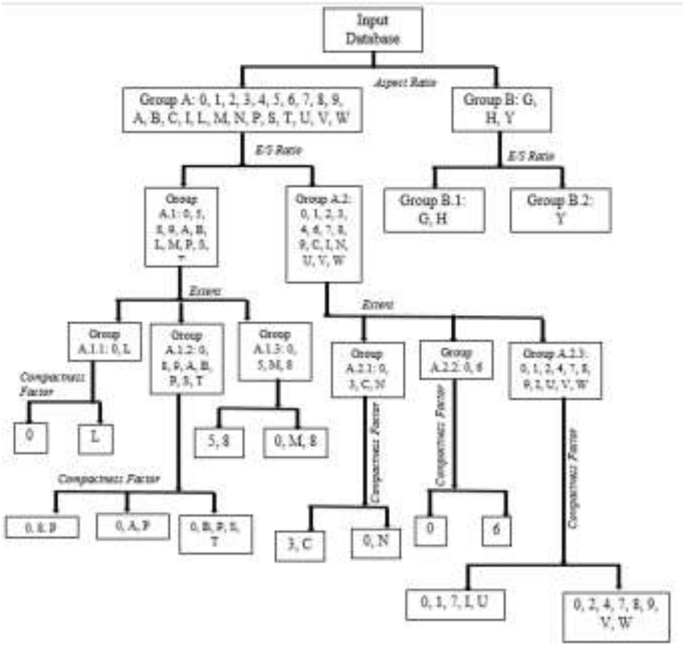

**Fig.8:** Classification based on appearance based model

D. **Template Matching**

Cross correlation is the process in which we compare the pixels of two different boudary-distance vectors and comparison of the result is shown in the form of graph.. Boundary-distance vector is obtained by calculating the distance between the centroid and boundary pixels. These distances are known as *Euclidian Distances*[4] shown in Fig 9(a). In the template matching, we exploit the symmetricity of the graph obtained. Graph would be symmetric only if it is compared with the image of same symbol. Fig.9(b) shows the symmetric and asymmetric curves.

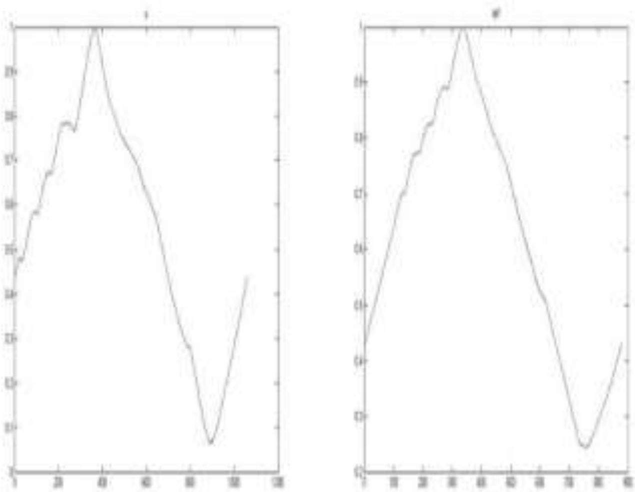

**Fig.9**: (a) Euclidian distance measures of a same symbol from different databses

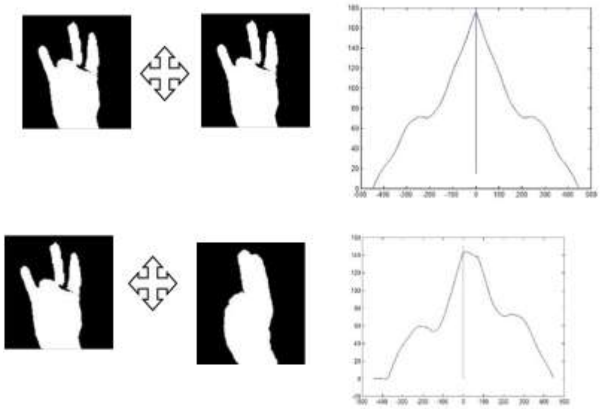

**Fig.9** (b). Symmetric and Asymmetric curves

### E. HOG Features

Histogram of Oriented Gradients (HOG) are feature descriptorscounts occurrences of gradient orientation in localized portions of an image.This features exploit the similarity property of the objects.The ideais that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions.HOG feature takes a gray scale or colored image as an inputand it gives a feature vector of dimension 1×N, where N is determined by cell and size of the image as shown in Fig.10.

After determining the feature vector, it calculate the value of correlation coefficient between the two compared featured vectors. The value is near to unity if both vectors are highly correlated, and it decreases as the correlation decreases.[5][6]

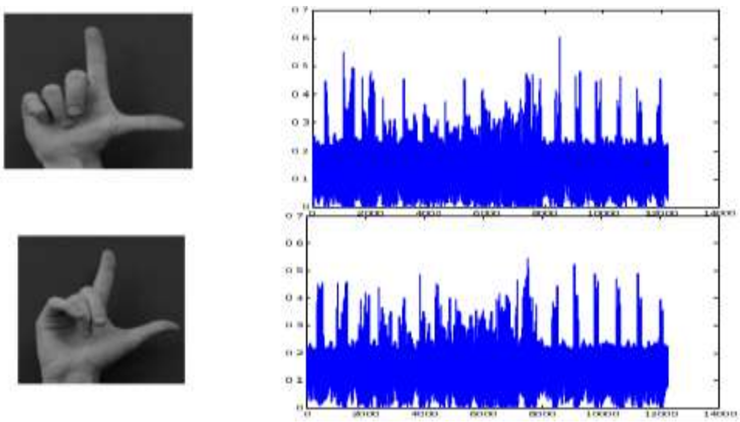

**Fig.10:** Gesture and its feature vector

### F. Angle Information

The remaining symbols are classified using peak information, number of peaks in the image and the location of peaks in the image as described the same in[4]

This feature is also most useful because we have used it in three different ways to classify the remaining symbols:

**1. Based on the number of peaks:** The number of peaks which means number of fingers in the image was counted in this type of classification.

**2. Based on the angular location of peaks:** When the number of peaks are same in any two images, it can be classified using this feature, in which we have calculated the angular position of the peaks.

**3. Angular location of threshold crossing:** In this feature, the height of the peak was calculated using the appropriate threshold.

Euclidian distances and the angle information were determined using the graphical plot of a particular gesture and necessary parameter is calculated.

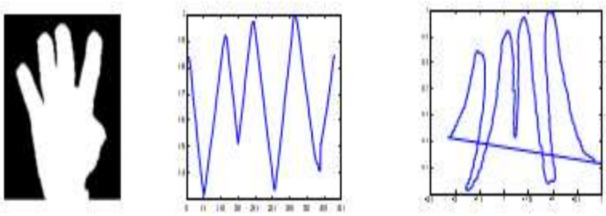

**Fig.10:** Euclidian Distances and Angle Information curves

Based on the position of the peak, we have distinguished few gestures. It is useful when the total number of peaks of Euclidian distances of two different images are same. It is shown in fig.12

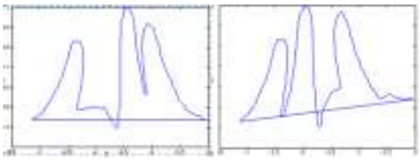

**Fig.11:** Peaks at different position

Apart from the peak position, classification of gestures is also done by quantizing its peak heights. Folding and straight finger have different heights to classify.

### G. Classifier

After classifying signs based on different features, task left is to develop a single platform on which all these operations can be carried out. Classifier is the name given to the module which contains every required process to execute the task.

Next section includes the results and the timing analysis of the designed algorithm.

## IV. RESULTS

The proposed algorithm was tested on three different databases. It is to note that the database images were having uniform background and all the test images were having the same dimension. It is to note that random images were taken from the databases to check the reliability of the algorithm.

The test result for the three database images and their recognition rate is specified in the Table 1.

**Table I**

| Input Gesture symbol | Test images taken | Successful Cases | Recognition (%) |
|---|---|---|---|
| 0 | 3 | 2 | 66 |
| 1 | 3 | 3 | 100 |
| 2 | 3 | 3 | 100 |
| 3 | 3 | 3 | 100 |
| 4 | 3 | 3 | 100 |
| 5 | 3 | 3 | 100 |
| 6 | 3 | 2 | 66 |
| 7 | 3 | 2 | 66 |
| 8 | 3 | 2 | 66 |
| 9 | 3 | 2 | 66 |
| A | 3 | 3 | 100 |
| B | 3 | 3 | 100 |
| C | 3 | 2 | 66 |
| G | 3 | 3 | 100 |
| H | 3 | 3 | 100 |
| I | 3 | 2 | 66 |
| K | 3 | 2 | 66 |
| L | 3 | 3 | 100 |
| M | 3 | 2 | 66 |

| | | | |
|---|---|---|---|
| N | 3 | 2 | 66 |
| P | 3 | 3 | 100 |
| S | 3 | 2 | 66 |
| T | 3 | 2 | 66 |
| U | 3 | 3 | 100 |
| V | 3 | 3 | 100 |
| W | 3 | 3 | 100 |
| Y | 3 | 3 | 100 |
| **ALL** | **81** | **69** | **85.18%** |

This table indicates that out of 81 database images, 69 were identified correctly with the success rate of around **85.18%**. 12 images failed to be recognized by this algorithm.

The timing analysis of the classifier is shown in the pie chart in Fig.13 which shows the total time taken for the execution and the time taken by the individual functions.

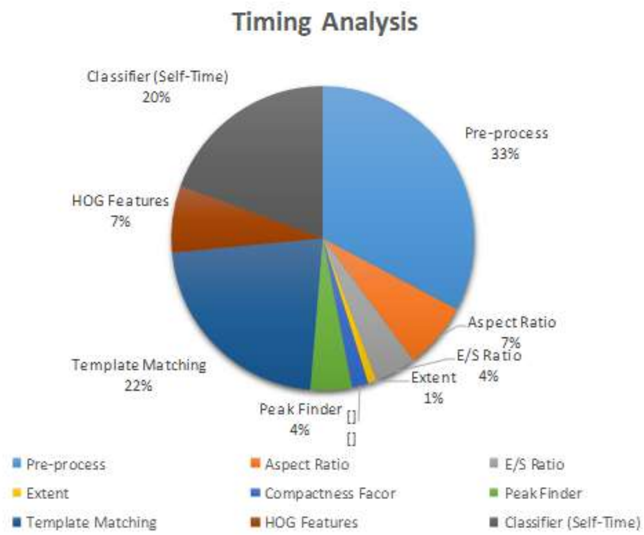


**Fig.12:** Timing Analysis of execution of each process

Average time of execution was around 2.408 milliseconds.

## V. CONCLUSION

Proposed algorithm is robust and efficient which can be used for the mentioned purpose. But in this work, controlled background is considered i.e. only single color is there in background, but for practical applications there can be uncontrolled background in the images. From just static alphabets, the project could well be extended to word recognition. Word recognition sets us the challenging task of manipulating videos, which is motion of the hands, needs to be considered as well. Since same symbol/ sign could mean different things in different context, the next task would be to add context and predictive sentence formation; which would require induction of artificial intelligence.

After applying this algorithm on few databases, we confirmed seven best possible features to classify all the images. The algorithm was tested on the free available database images and it showed around 86% success rate. The algorithm is modified to get it suited for real time applications also. The foremost advantage of this algorithm than the previous one is its accuracy; it can more accurately recognize the gestures with scaling and rotational invariability

## REFERENCES

[1]. Yanmin Zhu, Zhibo Yang and Bo Yuan, “Vision Based Hand Gesture Recognition”, International Conference on Service Science, pp. 260-265, 2013.

[2]. Pragati Garg, Naveen Aggarwal and Sanjeev Sofat, “Vision Based Hand Gesture Recognition System”, World Academy of Science, Engineering and Technology 49, pp. 972-976, 2009.

[3]. D.K.Vishwakarma and Rajiv Kapoor, “Simple and intelligent system to recognize the expression of speech-disabled person”, IEEE Proceedings of 4th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, December, 2012.

[4]. Anupam Agrawal, Rohit Raj and Shubha Porwal, “Vision-based Multimodal Human-Computer Interaction using Hand and Head Gestures”, IEEE Conference on Information and Communication Technologies (ICT 2013), pp. 1288-1291, 2013.

[5]. Navneet Dalal and Bill Triggs, “Histograms of Oriented Gradients for Human Detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), 2005.

[6]. Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng, “Fast Human Detection Using a Cascade of Histograms of Oriented Gradients”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.

[7]. Haiying Guan, Rogerio S. Feris, and Matthew Turk, “The isometric self-organizing map for 3d hand pose estimation,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 263–268.

[8]. Makoto Kato, Yen-Wei Chen, and Gang Xu, “Articulated hand tracking by pca-ica approach,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 329 – 334.

[9]. W. T. Freeman and C. Weissman, “Television control by hand gestures,” in Proceedings of International Workshop on Automatic Face and Gesture Recognition. Zurich, Switzerland, June 1995, pp. 197–183.

[10]. Y. Cui and J. Weng, “View-based hand segmentation and hand sequence recognition with complex backgrounds,” in Proceedings of 13th ICPR. Vienna, Austria, Aug. 1996, vol. 3, pp. 617– 621.

[11]. Just A., Rodriguez Y., and Marcel S., “Hand posture classification and recognition using the modified census transform,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 351–356.

[12]. J. Triesch and C. von der Malsburg, “Robust classification of hand posture against complex background,” in Proceedings of Int. Conf. on Face and Gesture Recognition. Killington, Vermont, Apr. 1996, pp. 170–175.

[13]. Lars Bretzner, Ivan Laptev, and Tony Lindeberg, “Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Washington D.C., May 2002, pp. 423–428.