International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 5, Issue 7 [Aug. 2016] PP: 21-30

Real Time Hand Gesture Recognition

Alhussain Akoum[1]
[1]*Department of CCNE, LEBANESE University, Institute of Technology, Saida*

***Abstract**:* In a general overview, signed language is a technique used for communicational purposes by deaf people. The American Sign Language (ASL) is a set of 26 meaningful hand gesture, each one representing a letter of the alphabet. In this report we will describe the algorithm for hand gesture recognition we developed based on Morphological analysis of the shape of the hand like the number of fingers, the angle between the finger, the number of peaks, the roundness and the orientation of the hand. The variation of these parameters optimized the results and delivered better classification criteria.

Keywords: *Camera, Hand gesture, image matching, image processing, ASL Language, Finger Detection, Roundness, gesture analysis, Orientation, recognition***.**

I. INTRODUCTION

Gesture recognition has been always a relatively fearful subject that is adherent to the individual on both academic and demonstrative levels. The core objective of this system is to produce a method which can identify detailed humanoid nods and use them to either deliver ones thoughts and feelings, or for device control. This system will stand as an effective replacement for speech, enhancing the individual's ability to express and intermingle in society. In this paper, we will discuss the different steps used to input, recognize and analyze the hand gestures, transforming it to both written words and audible speech. Each step is an independent algorithm that has its unique variables and conditions. Gesture is motion of body that contains information. [1] The straightforward purpose of a gesture is to express gen or interrelate with the surroundings. Motionless gestures are those that undertake a precise posted stance. Active contains a gesture movement that is distinct. Based on the locality of initiation of sign in the body, it can be considered a hand, an arm, a head or a face gesticulation. Natural HGR is one of the very active research areas in the Computer Vision field. It provides the easiness to interact with machines without using any extra device and if the users don't have much technical knowledge about the system, they still will be able to use the system with their normal hands. Gestures communicate the meaning of statement said by the human being. They come naturally with the words to help the receiver to understand the communication. It allows individuals to communicate feelings and thoughts with different emotions with words or without words. [4]

In our due time, software for sign language recognition is very imperative and is receiving great attention. Such software not only enhances communication between talking people and silent people, but also provides deaf people the ability to interact quickly and professionally with computers and machines using nothing but there hands. American Sign Language (ASL) is a complete system that is considered both simple and complex. It uses 26 different hand signs each indicating a letter. ASL is more than 200 years old. It is the preferable language of 500,000 deaf throughout the United States which rated it as the fourth most-used language.

This language is gaining attractiveness since it supports and enhances communication with an automated system or human located at a distance. Once the user finishes the gesture the system need to be capable of identifying it instantly. This is known as 'Gesture Recognition'. The target of this effort is to construct a system which can classify particular hand gestures and extract the corresponding literatures. This system dynamic, real-time and is based on the American Sign Language alphabets.

One of the ways to give signal to computer vision devices is by using hand gesture. More specifically hand gesture is used as the signal or input modality to the computer. Certain signal can be recognized by computer as an input of what computer should do. These will benefits the vision device can sense it. These make computer user easier than using the keyboard or mouse. The future computer or laptop may eliminate the use of keyboard and mouse by substituting with a vision-based interpretation devices. Interaction between humans comes from different sensory modes like gesture, speech, facial and body expressions [2]. The main advantage of using hand gestures is to interact with computer as a non-contact human computer input modality. The state of art of human computer interaction presents the facts that for controlling the computer processes gestures of various types of hand movements have been used .The present research effort defines an environment where a number of challenges have been considered for obtaining the hand gesture recognition techniques in the virtual environment. Being an interesting part of the Human computer interaction hand gesture recognition needs to be robust for real life applications, but complex structure of human hand presents a series of challenges for being tracked and interpreted. Other than the gesture complexities like variability and flexibility of structure of hand other challenges include the shape of gestures, real time application issues, presence of background noise and variations in illumination conditions. The specifications also involve accuracy of detection and recognition for The present research effort has a goal of developing an application using vision based hand gestures for manipulation of objects in virtual environment. Our application presents a more effective and user friendly methods of human computer interaction intelligently with the usage of hand gestures. Functions of mouse like controlling of movement of virtual object have been replaced by hand gestures. The complexity involved is with the detection and recognition phases of the simulated virtual application. The challenges encountered are noisy environment which creates a big impingement on the detection and recognition performance of human hand gestures. The application has been designed to be cost effective and uses low cost input tools like webcam for capturing hand as input. Manipulation of virtual objects has been done through modelling of some predefined command based hand gestures (Fig. 1).

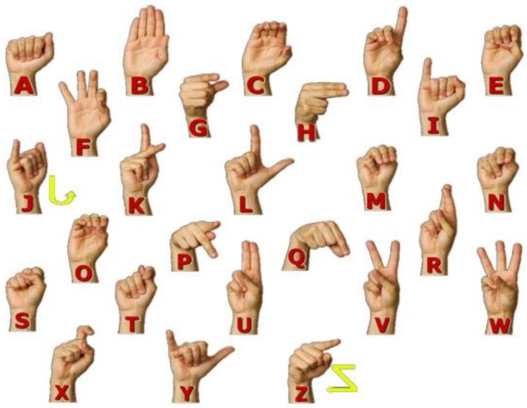


Fig 1. American Sign Language (ASL)

I. ALGORITHM WORK FLOW

The system that we worked on proved to follow a different approach to realize a more precise result. Though our method is more divergent and somehow complicated in terms of code, it is as simple as moving the hand in terms of usage. There are different parameters used in building our input data images (Fig. 2).

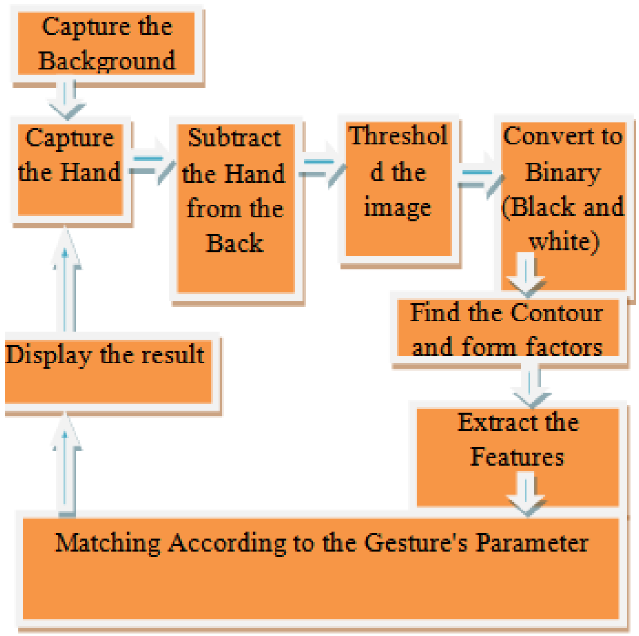


Fig 2. Algorithm Work Flow

1.1 Capturing

The By camera we are referring to images captured from the digital camera. It is VERY IMPORTANT to mention that the hardware we are working on, especially the camera, are neither professional nor has any profitable aims. Thus, facing certain obstacles and errors is inescapable. Such errors are not the results of the code but the technical insufficiency.

The Background should be clear and uniform (Fig. 3). The Hand must be fixed at a certain action in order to let the camera read this action, the video is a real time image (live).

Fig 3. Original image taken by the camera live.

1.2 Background Extraction

Background extraction is an important part of moving object detection algorithms that are very useful in surveillance systems. Moving object detection algorithm will be simple by background subtraction when a clean background image is available (Fig. 4).

Capture an image with the same back ground but without the hand with flexible conditions (light, Pixels…).

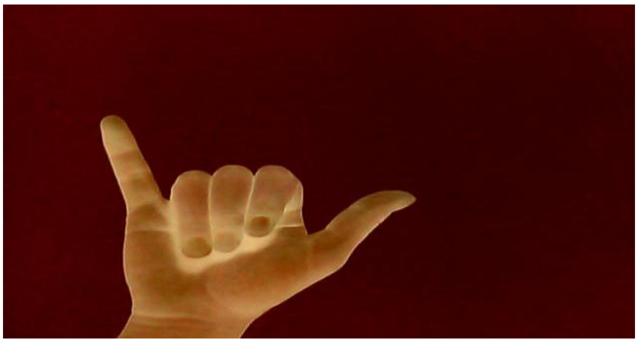


Fig 4. Background Removed

This is the method of Subtraction the background during the input frame sequence. The main challenges in moving object detection is extraction a clean background and its updating. Thus, subtracting the background after capturing the image would preserve only the hand.

1.3 Processing

Mechanism functions depending on the number of layers, RGB Euclidian threshold and fraction of observed color accumulating in the background.

We change the image type from ‘RGB’ to 'Gray' then to ‘Binary’ and applies filtering at different levels. Converting images to binary type is done by replacing all pixels according to the specified luminance with either white (logical 1) if the pixel is equal or greater that the level or black (logical 0) otherwise. Specified level should belong to the range [0, 1]. This level differs from one image to another. Though it can be manually assigned, using the function ‘Gray thresh’ to compute it gives better results. Other functions for processing: imerode will enlarge the black areas and eat away at the white areas. imdilate will enlarge/smooth the white areas and fill in black areas near borders/perimeters. imfill fills in black islands in a white blob.

The next step is removing small connected components and objects from binary image with bwareopen. Those objects have fewer pixels than the specified threshold. Removing the noise in the image is one of the most important and most difficult of the pre-handling techniques. This noise is designated as an unsystematic discrepancy of brightness or color gen through the image’s background (Fig. 5).

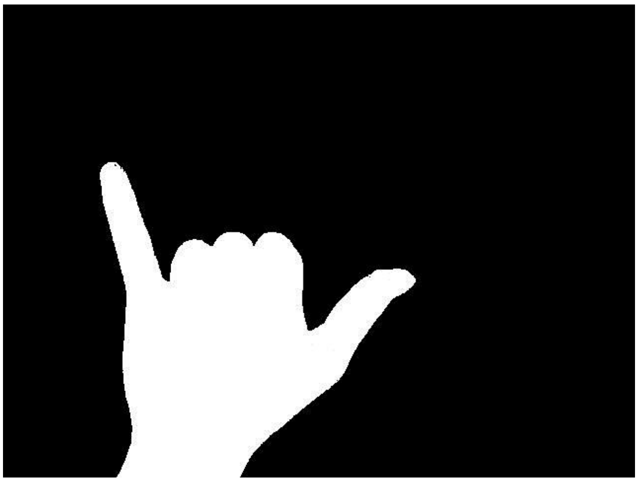


Fig 5. Processed binary image

1.4 Features and Form Factor

In order to have an accurate recognition for each gesture we included many factors and parameters to be extracted from the input image and compared with the database:

1.4.1 Number of fingers

This is a key factor for the classification that helps to widely distinct gestures and performs a more specific matching process (Fig. 6).

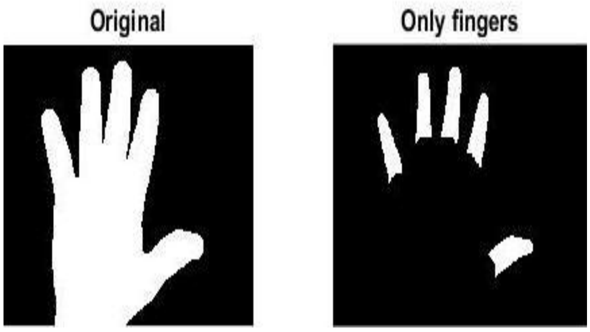


Fig 6. Fingers extracted from original input

The basic flow of this function is the next:

- Find the nearest point from centroid of extracted blob (which we might safely assume to be palm)
- The distance from the point can safely be assumed to approximate palm radius.
- With a disc element of a nearly this distance, erode the image.
- Subtract the resultant image2 from image1
- Do a few more morphological operations, to make fingers significant in appearance.

1.4.2 Orientation of the gesture

By definition, the Orientation in a black and white image is the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region. The value is in degrees, ranging from -90 to 90 degrees. This figure illustrates the axes and orientation of the ellipse (Fig. 7). The left side of the figure shows an image region and its corresponding ellipse. The right side shows the same ellipse with the solid blue lines representing the axes, the red dots is the foci, and the orientation is the angle between the horizontal dotted line and the major axis (Fig. 8).

Fig 7. Shape orientation

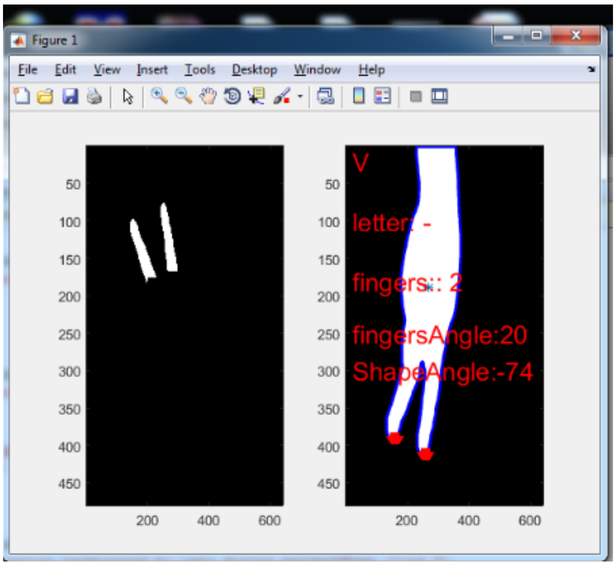


Fig 8. Gesture recognition using different parameters

1.4.3 Angle between fingers

This factor can differentiate between gestures with same number of fingers for example V and K have both 2 fingers but the angle between them is different. For the calculation of the angle the same idea of the shape orientation is used but in this case the measurement is relatively between the axis of the two ellipses (Fig. 9).

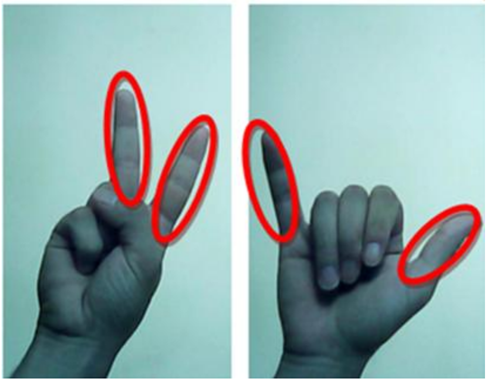
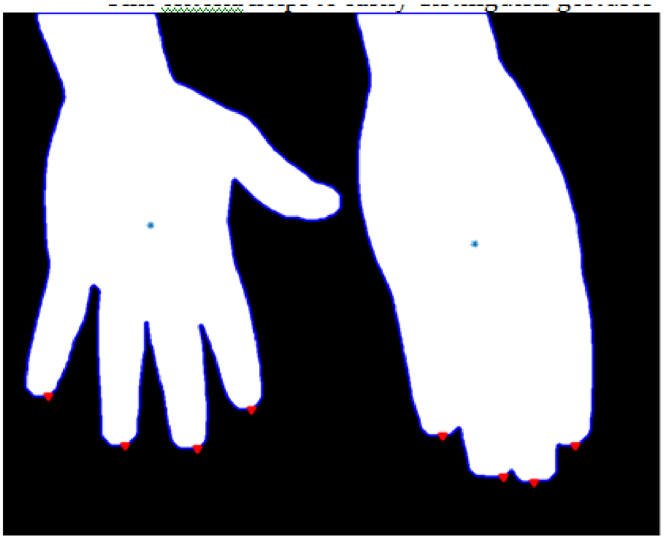
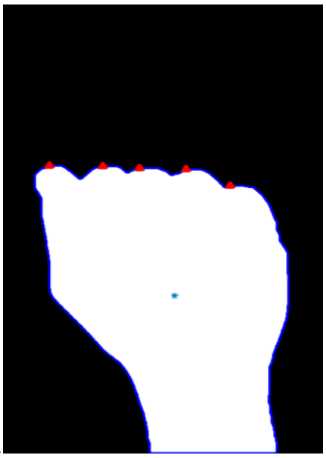


Fig 9. Ellipses containing the fingers

1.4.4 Number of PeaksDetecting the peaks values in the boundaries of the shape (Contour) Not only that this method detects the tip of a finger but also the bones structure (root of the finger) even if the gesture is done with

1.4.5





1.4.6 no fingers like the letter A (Fig. 10).

Fig 10. Peaks detection

1.4.7 Roundness

Roundness is useful, for an object of same shape ratio, regardless of size the roundness value remains the same. For instance, a circle with radius 5pixels will have the same roundness as a circle with radius 100 pixels. It is a measure of how round an object is.

$$\mathrm{R} = \frac{P^2}{4\pi A} \qquad (1)$$

This criteria helps to easily distinguish gestures with similar Area but different Perimeter (Fig. 11).

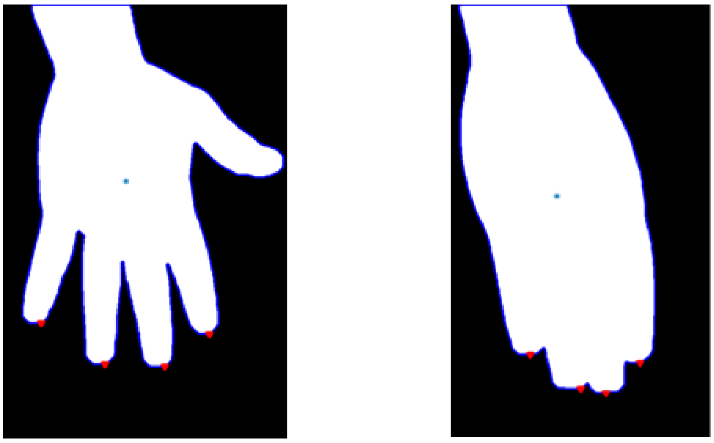


Fig 11. Images with different Roundness ratio

1.4.8 ConvexHull

The Convex Hull is a Polygon drawn by connecting the extreme points of the shape and thus containing the shape, we can use its Area, Perimeter and others as parameters (Fig. 12).

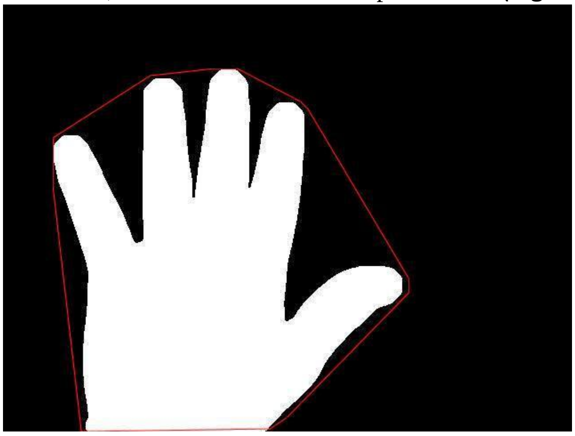


Fig 12. Convex hull Polygon

B. Feature Extraction

Now after we successfully detected all the features, we store them in a vector that will represent the gesture in the matching process.

II. IMAGE MATCHING, CREATING AND LOADING THE DATABASE

Database technology has matured with the development of relational databases, object-relational databases, and object-oriented databases. Extensions to these databases have been developed to handle nontraditional data types, such as images, video, audio, and maps. The core functionalities of classical databases, however, are tailored toward simple data types and do not extend gracefully to nonstructural information. Digital images have a predominant position among multimedia data types. Unlike video and audio, that is mostly used by the entertainment and news industry.

Images are vital to a comprehensive selection of fields including Communications and programming. Digital images play a valued part in abundant deeds, such as object and face recognition. Moreover, the utmost imperative advantage of storing images in a database is the standardizing of the data; i.e. any change applied on the images during the code would not affect the original values. When needing to work with a large amount of images, especially for matching purposes, rather than calling every image individually, gathering them in a database not only organize the data, but also guarantee a comparability using several methods and tools with normalized results.

Our database contains a normalized set of parameters for each letter of the American Sign Language samples stored in a MAT-file. It is important to mention that having a big database means more accuracy, which is likely to increase the recognition percentage. However, working on software, we must take into consideration the quality of the image as well as the overall size of the program. Since all what we care about in the input image is the hand, we resized the database to fit the description, and thus dramatically decreasing the size of the overall code. Note that later in the code, all images would be changed form 'RGB' to 'Binary' which reduces the image's size while preserving all the information we require from the gesture.

Run the matching algorithms on the processed image attempting to find the closest match. The models are in the form of pure noiseless images stored in a database.

Reduces the image's size while preserving all the information we require from the gesture. And for a better and more focused comparing process, on each iteration of the comparison algorithm, the vector of the input gesture is multiplied by a mask that corresponds specifically for the letter and which focuses on the true representative parameters and features of the gestures.

TABLE I

REPRESENTATIVE VECTOR FOR EACH LETTER IN THE DATABASE

	Roundness	**Nb. peaks**	**Shape angle**	**Fingers angle**	**Nb. fingers**
A	0.865264	4	-49.268684	57.8783021	0
B	0.7555149	3	-79.570192	67.7754138	0
C	0.413633	1	-33.707527	58.0639978	2
D	0.6517087	3	-71.93047	57.4728436	1
E	0.8649662	2	-77.558819	59.5670578	0
F	0.3169257	4	65.3145688	94.2904305	3
G	0.6068262	1	76.8891656	60.6484703	1
H	0.4534483	2	74.7197938	60.1078725	1
I	0.647691	3	-81.788867	65.2299398	1
J	0.6033184	2	-35.563829	48.8616123	1
K	0.4571964	2	-73.035929	60.6248264	2
L	0.4755841	3	-53.885198	72.3122072	2
M	0.7948631	2	-67.609685	60.6649482	0
N	0.7642405	2	-66.518962	59.009551	0
O	0.8001022	1	-66.164645	48.1615119	0
P	0.5875091	1	-32.778766	56.9347695	1
Q	0.4954117	1	35.6494443	76.5431985	1
R	0.6428398	2	-71.404863	53.991606	1
S	0.2863471	4	-86.450353	79.177963	5
T	0.5730257	2	-58.907921	71.8649363	1
U	0.6417142	2	-71.448625	52.0090728	1
V	0.4674647	2	-76.191387	57.3391711	2

W	0.4237482	3	-79.353299	57.4803732	3
X	0.6012657	1	-57.506285	49.6189169	1
Y	0.4494203	5	85.2320406	66.29207	2
Z	0.6159197	1	81.2627134	54.4126649	1

III. MATCHING OPTIMIZATION-TREE ALGORITHM

In order to optimize the matching process we developed a matching technique to classify results and accelerate the performance, which is based on a tree-like algorithm and is presented as following:

- Sort the parameters
- Classify gestures into sub-groups
- Each group represented by a parameter

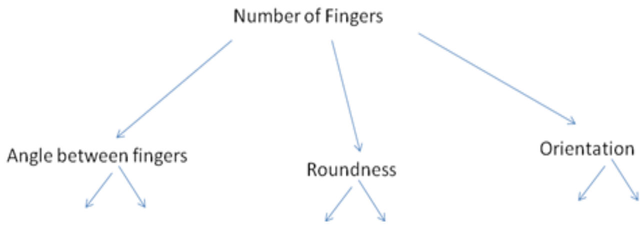


Fig 12. Classify gestures

A. *Results*

After taking 10 samples of each gesture, below is a chart representing the percentage of recognition for each of the letter (Fig 13).

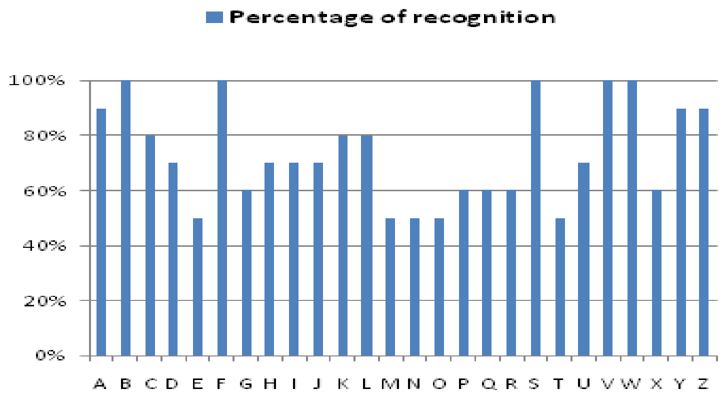


Fig 13. Percentage of recognition for each letter

IV. CONCLUSION

In present a number of facilities and various modes for providing input to any environment application are available. It is though unfortunate that with the ever increasing smart environments and corresponding input technologies still not many applications are available which are controlled using current and smart facility of providing input which is by hand gesture. The most important advantage of the usage of hand gesture based input modes is that using this method the user can interact with the application from a distance without using the keyboard or mouse. The application of manipulating objects through hand gestures in virtual environment is being proposed and implemented provides a suitable efficient and user friendly human computer interface. With the help of this application the user can interact with the virtual objects using hand gesture instead of any other physical input devices .As the application provides the flexibility to the users and specifically physically challenged users to define the gesture according to their feasibility and ease of use.

The present application though seems to be feasible and more user friendly in comparison to the traditional input modes but is somewhat less robust in recognition phase. An attempt to make the input modes less constraints dependent for the users hand gestures has been preferred. Another important aspect for the related development could be Hand extraction in a noisy environment.

REFERENCES

[1] Conic, N., Cerseato, P., De & Natale, F. G. B., “Natural Human- Machine Interface using an Interactive Virtual Blackboard”, In Proceeding of ICIP 2007, pp.181-184.

[2] Ismail, N. A., & O’Brien, A.,“ Enabling Multimodal Interaction in Web-Based Personal Digital Photo Browsing”, Proceedings of the International Conference on Computer and Communication Engineering , Kuala Lumpur, Malaysia, May 2008, pp. 907-910.

[3] [3] Pang, Y. Y., Ismail, N. A., & Gilbert, P. L. S.,“ A Real Time Vision-Based Hand Gesture Interaction”, Fourth Asia International Conference on Mathematical Analytical Modelling and Computer Simulation, 2010, pp. 237-242.

[4] Kortum, P., “HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces” Morgan Kaufmann Publishers, 2008, pp. 75-106.

[5] Viola & Jones, “Rapid object detection using boosted cascade of simple features", In Proceedings of Computer Vision and Pattern Recognition, pp. I-511 - I-518, 2001.

[6] Chen, Q., Coredea, M. D., Petriu, E. M., Varkony, A. R., Koczy, I., & Whalen, T.E., “ Human Computer Interaction for Smart Applications Using Hand Gesture and Facial Expressions,” International Journal of Advanced Media and Communication, vol. 3c.1/2, 2009, pp. 95-109.

[7] Jain, G. “Vision-Based Hand Gesture Pose Estimation for Mobile Devices”, University of Toronto, 2009.

[8] Pavlovic. V., Sharma, R., & Huang, T.S., “Visual interpretation of hand gestures for humancomputer interaction: A review.” IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 1997, 7(19):pp. 677–695.

[9] Marcel, S., Bernier, O., Viallet, J. E., & Collobert, D., “Hand Gesture Recognition using Input- Output Hidden Markov Models.” In Proc. of the FG’2000 Conference on Automatic Face and Gesture Recognition.

[10] Rautaray, S.S., & Agrawal, A., “A Novel Human Computer Interface Based On Hand Gesture Recognition Using Computer Vision Techniques”, In Proceedings of ACM IITM’10, 2010, pp.292-296.

[11] Aran, O., Ari, I., Benoit, F., C Campr, A., Carrillo, A. H., Fanard, Akarun, L., Caplier, a., Rombaut, M., & Sankuru, B., (2006) “ Sign Language Tutoring Tool”, eNTERFACE 2006, The Summer Workshop on Multimodal Interfaces, Croatia.

[12] Liu, N., & Lovell, B., “Mmx-accelerated realtime hand tracking system” In Proceedings ofIVCNZ, 2001.

[13] F. Chen, C. Fu, & C. Huang, “Hand gesture recognition using a real-time tracking method andhidden Markov models” Image and Vision Computing, , 2003, pp. 745-758.

[14] Lee, C. S., Ghyme, S. W., Park, C. J., Wohn, K., “The Control of avatar motion using hand gesture”, In Proceeding of Virtual Reality Software and technology (VRST) , 1998, pp. 59-65.

[15] Ahn, S. C., Lee, T. S., Kim, I. J., Kwon, Y. M., & Kim, H. G., “Computer Vision-Based Interactive Presentation System,” Proceedings of Asian Conference for Computer Vision, 2004.