Gesture Recognition based 3D HCI

Ramandeep Kaur[1], Vijay Laxmi[2]

[1]M.Tech, Department of Computer Science and Engineering, Guru Kashi University, Talwandi Sabo, Punjab, India

[2]Associate Professor, Department of Computer Science and Engineering, Guru Kashi University, Talwandi Sabo, Punjab, India

Abstract: *Digital Image Processing is one of the basic and important tools in the image processing and computer vision. Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with the intent of conveying meaningful information or interacting with the environment. Gesture recognition has wide-ranging applications. Creating a virtual human computer interaction device such as mouse or keyboard using a webcam and computer vision techniques can be an alternative way for the touch screen. Many researchers in the human computer interaction and robotics fields have tried to control mouse movement using video devices. This kind of human-machine interfaces would allow a user to control a wide variety of devices through hand gestures. This paper proposes the controlling of an object in virtual 3D environment using the data fetched fromextracting and tracking gestures from live images.Cascade object detector will be used to detect the gestures. Cascade object detector can be trained to detect a variety of objects. Based on the detected hand gestures, we will control different operations on the object in 3D environment.*

Keywords: Gesture, Human Machine Interface (HMI), Human Computer Interface (HCI), 3-Dimension, recognition

1. Introduction

Since the computer technology continues to grow up, the importance of human computer interaction is enormously increasing. Nowadays most of the mobile devices are using a touch screen technology. However, this technology is still not cheap enough to be used in desktop systems.

Human computer interaction is the study of people, computer technology, and the ways these two influence each other [R]. However, in practice Jacko et al defines HCI as, "a discipline that involves the design, evaluation and implementation of interactive computing systems for human use. According to Yee[S], HCI refers to: "the way a human and computer communicate using set of physical and logical rules." Hence, Yee [S] refers to the way a person experiences computer systems. The HCI discipline has evolved mainly due to two driving forces namely: the evolving nature of computers, and the desire for more powerful and compelling user experience. HCI tends to be characterized by new means of interaction which are natural, adaptive, intuitive and unobtrusive and mimic human-human communication. The introduction of such interaction methods has benefited many users including those with motor impairments. Most importantly designed are interfaces with a range of inputs and outputs it provides.

1.1 Unimodal HCI Systems

As revealed earlier, an interface depends on number and variety of its inputs and outputs which are communication channels that enable users to interact with a computer. According to Alemzadeh et al each of the different channels is called a modality. A system that supports only one modality is called unimodal. Below are two categories of unimodal systems:
1. Vision -based
2. Audio-based

Vision based human computer interaction provides a broad range of input capabilities by employing computer vision techniques to process sensor data from one or more cameras in real-time. This is done to reliably estimate relevant visual information about the user. Vision based interaction has the ability to carry rich information in a non-intrusive manner [T].

Jadon et al [T] proposed two approaches which are commonly used to interpret gestures for HCI:
1) *Methods which use data gloves*: These methods use sensors attached to the glove that transduce finger inflexion to electrical signals for determining hand postures. However this approach forces a user to carry a lot of cables needed for connection to the computer. Hence it compromises the naturalness and easiness of using vision-based interfaces.
2) *Methods which are vision-based*: These methods are non-invasive and intuitive because they are based on the way human perceive information about their environment.

2. Proposed Technique

In our research we have proposed a system to extract and track gestures from live images and using the fetched data to create a virtual interface for 3D human computer interaction. We will use cascade object detector to detect the gestures. Cascade object detector can be trained to detect a variety of objects. Based on the detected hand gestures, we will control the object in virtual 3D environment.

2.1 Cascade Object Detector

The cascade object detector uses the Viola-Jones algorithm to detect people's faces, noses, eyes, mouth, or upper body. We can also use the *train Cascade Object Detector* function to train a custom classifier to use with the *vision. Cascade Object Detector* System object. The *vision. Cascade Object Detector* System object comes with several pretrained classifiers for detecting frontal faces, profile faces, noses, upperbody, and eyes. However, these classifiers may not be sufficient for a particular application. Computer Vision System Toolbox™ provides the *trainCascade Object Detector* function to train a custom classifier.
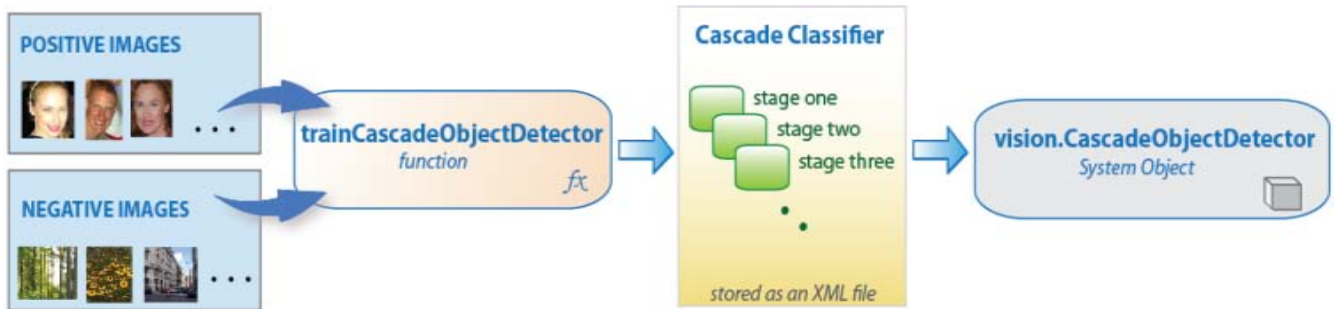
Paper ID: SUB14271

15

Figure 1: Overview of *trainCascadeObjectDetector*

2.2 Preview function

The preview function creates the preview of live video data. preview(obj) creates a Video Preview window that displays live video data for video input object obj. The window also displays the timestamp and video resolution of each frame, and the current status of obj (Fig. 2).
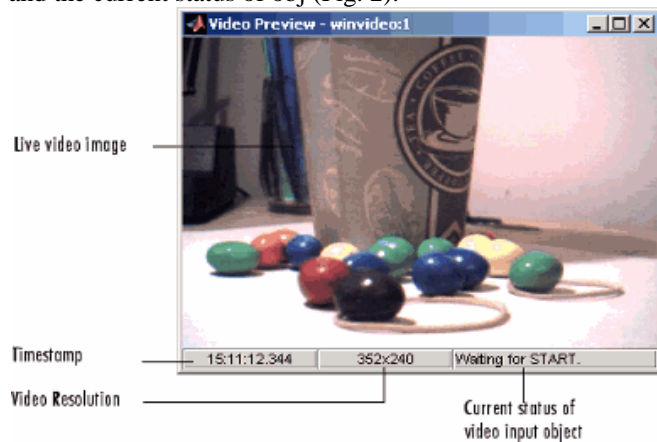

Figure 2: Components of Video Preview Window

2.2.1 Custom Update Function

preview creates application-defined data for the image object, *h image*, assigning it the name '*Update Preview Window Fcn*' and setting its value to an empty array ([]). You can configure the value of the '*Update Preview Window Fcn*' application data and retrieve its value using the MATLAB *setappdata* and *getappdata* functions, respectively. In this way, use of such persistent variables limits memory allocation. All the calls tographics can also be minimized by only updating the data, not the figure or the axes.

3. Proposed Procedure

We are implementing experimental research methodology by employing the procedure discussed below.

3.1 Defining Gestures

Gestures are meaningful body or body parts movements. We have used hand gesture in our research. These gestures are very simple hand postures to indicate different operations. Following are various gestures used in our research:

Figure 3: Various hand gestures used in the research

3.2 Recording Gesture Images in Digital Form

Gestures are recorded with a digital camera with plane backgrounds. A large set of small variation of each gestures are recorded to enable training of gestures for slightest orientations.

3.3 Extracting Gesture Information for Gesture Recognition

For extracting gesture information from the recorded data Cascade object trainer is used. For training we need a set of samples. There are two types of samples: negative and positive. Negative samples correspond to non-object images. Positive samples correspond to images with detected objects. Set of negative samples must be prepared manually, whereas set of positive samples is created using *training Image Labeler*.

3.4 Mapping Gestures with Operations

With the *trainCascadeObjectDetector* function we will map the pre-defined hand gestures with specific operations in 3D environment to be performed like translation, rotation and scaling.

3.5 Runtime gesture capturing from video device

Runtime gesture frame is captured using the preview function of video input class of MATLAB in RGB color space. Five frames are acquired per trigger and for each

frame acquired, a displace function is called which compares the gestures and executes the operation. The frame is passed to cascade object detector which searches the frame for each instance of the required gestures.

3.6 Simulating basic 3D operations with gestures

A basic algorithm is created to perform basic 3D operations i.e., move left, move right, rotate left, rotate right, scale up, scale down etc. For every frame captured the gesture is detected and its type is passed to a function which performs equivalent operations.

The following parameters have been defined to carry out research and evaluate the performance:

a) *True Detection Rate*:

$$TDR = \frac{\text{No. of true gestures detected}}{\text{Total no. of gestures detected}} * 100$$
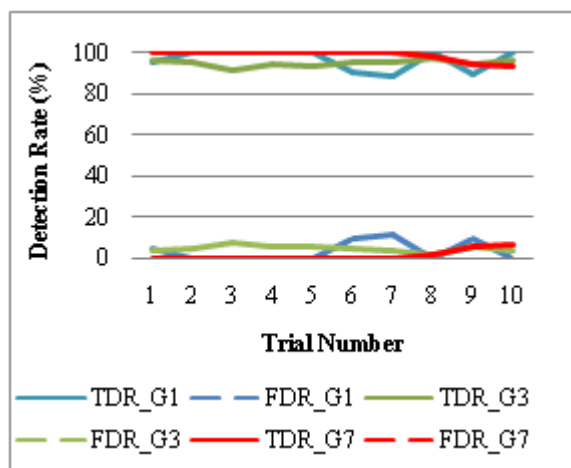
b) *False Detection Rate*:

$$FDR = \frac{\text{No. of false gestures detected}}{\text{Total no. of gestures detected}} * 100$$

4. Results

The experiments are executed in Matlab R2013a on a laptop with Intel Core 2 Duo processor @ 1.66 GHz and 4GB RAM. Images are captured in YUY2_320x240 format with Microsoft LifeCam, which is a simple webcam. The performance is evaluated based on the overall gesture recognition rate.

In total, 8 gestures have been trained. As Gesture 1 has been trained up to maximum number of stages, it will give most accurate results out of these gestures. Similarly Gesture 3 and 7 will also yield much accurate results but Gesture 6 and 8 will not be much better option. So, results are based on the performance of best trained gestures, i.e. gesture 1, 3 and 7.

The detection rates can be compared graphically as:

Graph 1: Comparison of detection rates of gesture 1,3 and 7 under 10 different trials

In short, the overall TDR and FDR of Gesture 1, 3 and 7 can be represented graphically as:
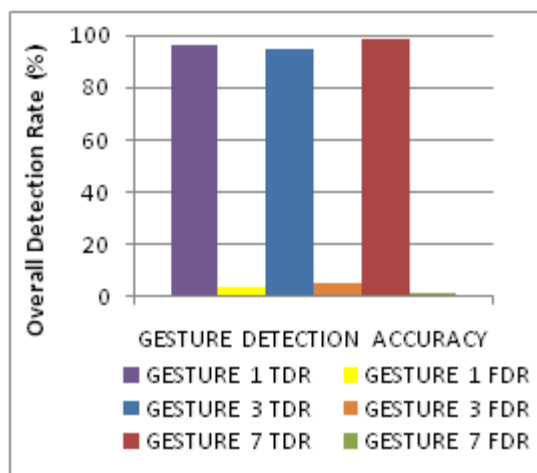
Graph 2: Comparison of overall detection rates of gesture 1,3 and 7

So, Overall accuracy
= (TDR (Gesture 1 + Gesture 3 + Gesture 7)) / 3
= (96.42115 + 95.1001+ 98.5807) / 3 = **96.7%**

This overall accuracy can be easily improved with more number of captured images of each gesture under different noise and environmental conditions.

There are many approaches to hand gesture recognition, and each approach has its strengths and weaknesses. We compare a gesture recognition efficiency using [E],[I] and the proposed gesture recognition system.

Amornched, J. et al proposed a simple recognition algorithm that uses three shape-based features of a hand to identify what gesture it is conveying. The algorithm takes an input image of a hand gesture and calculates three features of the image, two based on compactness, and one based on radial distance. The algorithm was tested on 200 hand images, and was able to successfully classify 182 images, or with an overall success rate of 91 percent.

The weakness of this method is the lack of a systematic approach to defining certain parameters. The threshold values for the three parameters were obtained empirically. Also the hand is partitioned into two halves at its geometric center, or centroid, by a vertical line parallel to the image edge. This makes the approach very restrictive in shape of hand which is quite hard to achieve.

Yang, C. et al (2012) proposed a gesture recognition system capable of providing a contactless controller via depth-based hand tracking. The images are captured with a Microsoft Kinect device, so the cost is quite high. Also the overall efficiency of this system is comparable to our proposed system even when the images in our research are captured by simple webcam.

The overall efficiency of the three systems mentioned above can be compared graphically as:
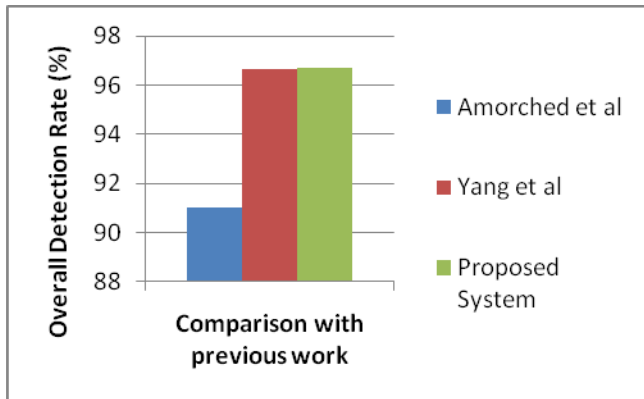
Graph 3: Comparison with previous techniques

Although our proposed system has comparable results with the system developed by Yang et al, but our system is purely based on simple webcam while the other is based on a costly device i.e., Microsoft Kinect.

5. Conclusion and Future Scope

Gesture based interfaces require a very fast real time detection system. The interfaces created with gesture based system are comparatively slower than hardware interfaces which makes it inapt to use for applications which demand real-time responses. Some of the areas where these interfaces can be deployed are:

- Smart TV
- Small games not requiring fast responses.

We proposed a gesture recognition system using information fetched from Cascade Object Detector and demonstrated its effectiveness in an application containing the object to be controlled. The result of hand gesture recognition shows robust performance, implying that the proposed system may be suitable to be applied to various contactless user interface systems. Also, the developed system is very cost effecting fulfilling the need for cheaper applications.

The efficiency of the system is confirmed through evaluation experiments. Furthermore, the developed system is made flexible enough to be easily manipulated. The user has been provided with flexibility in choosing any of the gestures for any feature which makes the system more reliable.
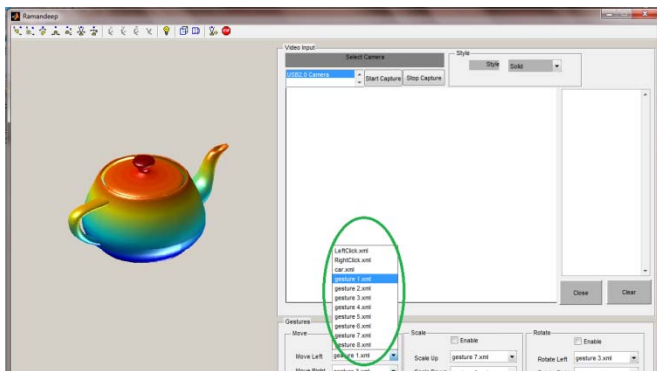

Figure 4: Flexibility in choosing a gesture

By incorporating further tracking algorithms, more 3D operations like slicing can be performed making the system more efficient. Also by adding more samples of each gesture

under different noise, light and other environment conditions the efficiency of the developed system can be increased.

References

[1] Hürst, W. and Wezel, C.V. (2012),*"Gesture-based interaction via finger tracking for mobile augmented reality"*, Springer.com.

[2] Argyros, A.A. and Lourakis, M.I.A. (2006),*"Vision-Based Interpretation of Hand Gestures for Remote Control of a Computer Mouse"*, Proceedings of the (European Conference on Computer Vision) ECCV, pp. 39-50.

[3] Grandhi, S.; Joue, G. and Mittelberg, I. (2012),*"What's in a Gesture: Exploring the potential of gesture-based interaction for low literacy user groups"*, Proceedings of the annual conference on Human factors in computing systems (CHI). ACM, Austin, Texas, USA.

[4] Bai, H.; Billinghurst, M.; Gao, L. and El-Sana, J. (2013),*"Markerless 3D Gesture-based Interaction for Handheld Augmented Reality Interfaces"*, Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, Australia.

[5] Yang, C.; Jang, Y.; Beh, J.; Han, D. and Hanseok, K.O. (2012),*"Gesture recognition using depth-based hand tracking for contactless controller application"*, IEEE International Conference on Consumer Electronics (ICCE),pp. 297-298.

[6] Igorevich, R.R.; Park, P.; Choi, J. and Min, D. (2012),*"iVision based Context-Aware Smart Home System"*, 1st Global Conference on Consumer Electronics (GCCE), IEEE, pp. 542 –546.

[7] Igorevich, R.R; Park, P.; Min, D.; Park,Y.; Choi, J. and Choi, E.(2010), *"Hand gesture recognition algorithm based on grayscale histogram of the image"*, 4th International Conference on Application of Information and Communication Technologies (AICT), IEEE, pp.1-4.

[8] Bray, M.; Koller-Meier, E. and Gool, M.V. (2004), *"Smart particle filtering for 3d hand tracking"*, Proceedings of the Sixth IEEE International Conference on Face and Gesture Recognition (FGR).

[9] Jinda-apiraksa, A.; Pongstiensak, W. and Kondo, T., (2010), *"A Simple Shape-based Approach to Hand Gesture Recognition"*, Proceedings of IEEE International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), Pathumthani, Thailand, pp. 851-855.

[10] Manchanda, K. and Bing, B. (2010), *"Advanced Mouse Pointer Control Using Trajectory-Based Gesture Recognition"*, Proceedings of IEEE southeast Conference on motion control, pp. 412-415.

[11] Ueda, Y. and Maeno, T., (2004), *"Development of a Mouse-Shaped Haptic Device with Multiple Finger Inputs"*, Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems , Sendai, Japan, pp. 2886-2891.

[12] Tsang, W.M. and Pun, K., (2005), *"A finger-tracking virtual mouse realized in an embedded system"*, Proceedings of IEEE International Symposium on Intelligent Signal Processing and Communication Systems, Hong Kong, pp. 781-784.

Paper ID: SUB14271

18

[13] An, J. and Hong, K.S., (2011), "*Finger Gesture-Based Mobile User Interface Using a Rear-facing Camera*", Proceedings of IEEE International Conference on Consumer Electronics (ICCE), pp. 303-304.

[14] Morris, T., and Chauhan, V., (2006), "*Facial feature tracking for cursor control*". J. Netw. Comput. Pp. 62–80.

[15] Erdem, E.; Yardimci, Y.; Atalay, V. and Cetin, A. E, (2002), "*Computer vision based mouse*", Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASS) IEEE, vol.4, no., pp.IV-4178.

[16] Lien, C., "*Portable Vision-Based HCI – A Real-time Hand Mouse System on Handheld Devices*", National Taiwan University, Computer Science and Information Engineering Department.

[17] Park, H., (2008), "*A Method for Controlling the Mouse Movement using a Real Time Camera*", Brown University, Providence, RI, USA, Department of Computer Science.

[18] Porta, M., (2002), "*Vision-based user interfaces: methods and applications*", International Journal of Human Computer Studies, vol. 57, issue 1, pp. 27-73.

[19] Yee, C.S.M., (2009), "*Advanced and natural interaction system for motion-impaired users*", PhD Thesis, Universitat de les IllesBalearsDepartament de CiuenciesMatemuatiquesiInformuatica.

[20] Murthy, G.R.S. and Jadon, R.S., (2009), "*A review of vision based hand gestures recognition*", International Journal of Information Technology and Knowledge Management, vol. 2, pp.405-410.

[21] Ozun, O.; Ozer, O.F.; Tutzel, C.O.; Atalay, V. and Cetin, A.E., (2001), "*Vision Based Single Stroke Character Recognition for Wearable Computing*", IEEE Intelligent Systems and Applications, pp.33-37.

[22] Yun Fu and Thomas S.H., (2007), "*Head Tracking driven Virtual Computer Mouse*", Proceedings of the Eighth IEEE Workshop on Applications of Computer Vision (WACV '07), pp.30.

[23] Malkewitz, R., (1998), "*Head Pointing and Speech Control as a Hands-free Interface to Desktop Computing*", Proceedings of the third international ACM conference on Assistive Technologies (ASSETS '98), pp.182-188.

[24] Jilin, T., Huang, T. and Tao, H., (2006), "*Face as Mouse through Visual Face Tracking*", Computer Vision and Image Understanding (CVIU) special issue on HCI, pp.35-40.

[25] Canada, C.; Gorodnichy, D.O.; Malik. S. and Roth, G, (2002), "*Nouse 'Use your Nose as a Mouse' – a new technology for hands free games and interfaces*", Proceedings of Vision Interface, pp. 354-361.

Author Profile

Er. RamandeepKaur, I have received my B.Tech degree in Computer Science & Engineering from GianiZail Singh College of Engineering and Technology, Bathinda (under Punjab Technical University, Jalandhar) in 2010 and pursuing M.Tech degree in Computer Science & Engineering from Guru Kashi University, Talwandi Sabo (Bathinda). My Research area is Digital Image Processing.

Dr Vijay Laxmireceived her B.Tech degree in computer science & Engineering from SLIET Longowal in 2003, and Ph.D degree in Computer Science & Engineering in year 2012. Her research area is Grid Computing & OCR. She has published 25 research papers in various National/International Conferences and Journals. At present, she is engaged in Guru Kashi University, Talwandi Sabo, and Punjab as Dy. Dean Research and an Associate Professor in Computer Science Engineering & Information Technology department.

Paper ID: SUB14271

19