ISBN 978-1-84626-xxx-x
Proceedings of 2011 International Conference on Optimization of the Robots and Manipulators (OPTIROB 2011)
Sinaia, *Romania, 26-28 Mai*, *2011*, *pp. xxx-xxx*

Real-time gesture recognition for controlling a virtual hand

Catalin Constantin Moldovan and Ionel Staretu[+]

"Transilvania" University of Brasov, Romania

Abstract. Object tracking in three dimensional environments is an area of research that has attracted a lot of attention lately, for its potential regarding the interaction between man and machine. Hand gesture detection and recognition, in real time, from video stream, plays a significant role in the human-computer interaction and, on the current digital image processing applications, this represent a difficult task. This paper aims to present a new method for human hand control in virtual environments, by eliminating the need of an external device currently used for hand motion capture and digitization. A first step in this direction would be the detection of human hand, followed by the detection of gestures and their use to control a virtual hand in a virtual environment.

Keywords: Grasping, Motion detection, Object detection, Object recognition, Virtual reality.

1. Introduction

Hand gestures represent a powerful way of communication between people. These are rooted deep in our subconscious, an example in this direction can be considered: some people tend to gesticulate even when they are talking on the phone [1].

In general we cannot define a natural communication between man and machine without using gestures. These can be associated, to some extend, with temporal and spatial structures; we can indicate directions or actions using gestures, etc.

The main obstacle in achieving a natural interaction between man and machine based on gestures is the lack of appropriate methods of recognition and interpretation of the gestures by the computer.

A wide range of devices were developed to capture and eventually to reduce the degrees of freedom total number. These devices can be classified into magnetically, mechanical, optical, acoustic and inertial tracking devices.

Using these types of devices presented above, in [1] it is presented a review of the interaction modes between man and machine using human gestures.

According to the publisher, methods of interaction can be divided into two main classes of methods:

- "Data-Gloved based"
- "Vision based"

Data-Gloves approaches are based on the use of sensors devices, which digitize the human hand and finger movements in input parameters for a virtual reality simulation system.

Vision based approaches use image capturing devices. In this way a more natural interaction is achieved. The main advantages of the "Vision based" approaches are:

- Elimination of physical contact with the device
- Reducing storage space of the devices oriented on "Data gloves" approaches

[+] Corresponding author. Tel.: + 40 743 366 266; fax: + 40 268 418 967.
E-mail address: catalin.moldovan@unitbv.ro.

- Longer usage time, by eliminating wear of mechanical parts.
- Lower costs
- On the same application multiple users can work simultaneously

The purpose of this research is to present a new and effective method to control a virtual hand using an innovative "Vision based" approach. Also the main advantages and disadvantages, using this method in comparison with the "Data gloves" methods currently used to control a virtual hand, will be listed.

2. Used Method

Based on the researches presented in [2] it can be observed that the human hand is an elaborated anatomical structure composed of several connected parts and joints that involves complex relations between them; giving a total of 29 degrees of freedom (DOF), 23 from the finger joints from over the palm, and the remaining six come from the general orientation of the hand measured from the centre of the palm.

Because it is desired that the total execution time to be minimized, it can be used in the first place a skin detection method, in the input image, to reduce the search space for hand detection.

Recently these approaches were mainly investigated. In [3] several techniques to detect skin pixels from an image are presented, starting with a direct approach, using colour spaces or skin modelling methods. From user observations, Vezhnevets deduced that these methods shown promising results if images with high resolution are used.

To detect skin images that are not initially manipulated, it can be used a dynamic method for fixing an initial limit, a variant of this method is presented in [4]. Basically, the main idea behind this method is to divide the images into two types of regions: skin and non-skin.

Once the image was divided into skin and non-skin regions, a method to search a human hand only in skin regions can be used.

A new version of AdaBoost algorithm, used to detect objects, is presented in [5]. AdaBoost is used in many empirical researches and has received a special attention in the artificial intelligence in the last few years. AdaBoost was initially designed, as it is shown in [6], for human face detection. This algorithm can be extended, as seen in [5] to detect any kind of object.

AdaBoost is the short name of Adaptive Boosting and it is a learning method formulated by Yoav Freund and Robert Schapire, it is a meta-algorithm that can be used with other learning algorithms to improve their overall performance.

The system introduces a set of Haar classifiers for the AdaBoost algorithm that was design to quickly eliminate all the non-faces and the speed of detection process is very high.

The algorithm can be adapted to use Haar classifiers to detect human hand and recognise the gestures. The classifier is a basic unit of object detection, being similar to a Haar function.

3. General structure of the system

3.1. Overview of the system

This research proposes a architecture for a system that detects human hand gestures in real time, and uses these gestures to control a virtual hand. The figure below shows an overview of the propose architecture for this system.

As it can be seen, the system consists of five modules, namely FrameGrabber, SkinSegmentation, PostureDetection, ControlHand, DisplayModule, summarized in the next paragraph.

In the "Vision Based" hand gestures detection systems, hand movements are recorded by a video camera. The algorithms of the system are executed separately on each frame from the entire video sequence.

To gain speed, before executing an algorithm, an initial filter that eliminates unnecessary data and highlights necessary data is applied.

3.2. Initialization

To implement the hand detection and gesture recognition module an image processing library called OpenCV was used.

OpenCV library [7] was used, in the first place, to calibrate Haar classifiers. This library represents a collection of subprograms and algorithms implementation generally used for image processing applications. These subprograms were rigorously tested and can be easily used by other applications without having to rewrite them from scratch.

OpenCV Library is free and can run on multiple operating systems (Windows / Linux / MacOS).

Using this library, HandTrainer application was created. This application uses the digital web camera to capture images, and creates a set of positive and negative sample images which are used to train Haar classifiers.

The application capture an image every 500 milliseconds and save that image on disk into a predefined folder. The application is used to capture the "positive" sample files which contains, at lease once, the object which it is desired to be recognized. To create positive samples, for each image, a text file, containing the coordinates of the rectangle that fits the object which is wanted to be detected, is required.

The training process requires lots of hardware resources and to train a classifier for a specific object detection several days of processing are required. The results represent a XML file in which all information separated in logical groups are gathered.

The training process need to be executed only once, after this, Haar classifiers can be used anytime to detect human hand from a video stream.

The XML files, representing Haar classifiers for human hand detection, trained in the previously step are used to control a virtual hand in the GraspIT application. This application represents a virtual simulator used to simulate human and robot grasping. [8].

GraspIT simulator allows a user to import a human hand and objects which can be gripped. Collision detection will be made in real time, leading to the finding contact points and associated forces between virtual human hand and a gripped object.

Because GraspIT simulator has applicability in various robotic research areas, especially in grasping simulation, it was requested by multiple design centers. ROBONAUT group from NASA, have expressed interest in using GraspIT application for robot grasping simulation in the future NASA missions.

3.3. Implementation

The **Frame Grabber** module is used to capture all images from a stream. This can represent a part from a video file, or directly the image stream from the webcam.

On this module, each frame, which later will be processed and contain human hand gesture, represent the output.

The SkinSegemention module uses as input data, the result image from the FrameGrabber module. On each image, The SkinSegmentation module will detect the regions that contain skin's color. The method used to detect skin images is presented in [4] at a conceptual level.

The output from this module represents the same image from FrameGrabber together with one array containing all pixels region where skin was detected.

PostureDetection module uses the array from SkinSegmentation in which Haar classifiers trained in the CreateSample module are used to detect the hand posture. Currently the system has been implemented to detect gestures for the following "Fig. 1":

Figure 1: Fist and open hand.

The ControlPosture module will generate input data for the control of a virtual hand in GraspIT simulator.

DisplayModule use the GraspIT application presented in [8] to simulate virtual hand “Fig. 2”.

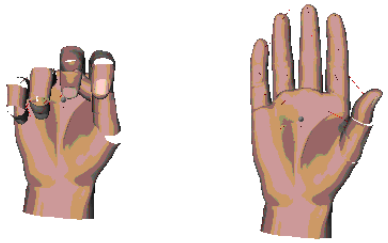

Figure 2: Simulated virtual hand. Fist and open hand.

4. Experimental results

In this section, an assessment and an analysis of the presented system are detailed. The performance of the system and of each module is analyzed.

In the following paragraph the robustness of the system is measured in different axis, compared with the method used to control a virtual hand using a CyberGlove device, which has 100% successful rate.

Method used to control a virtual hand with an external device will not be affected by changes in light condition, complex background, etc.

When the hand moves around the X and the Y axis the success rate do not vary. In the following graphic it is presented the robustness of the system when the hand position to evaluate moves along the axis "Z" “Fig. 3”:

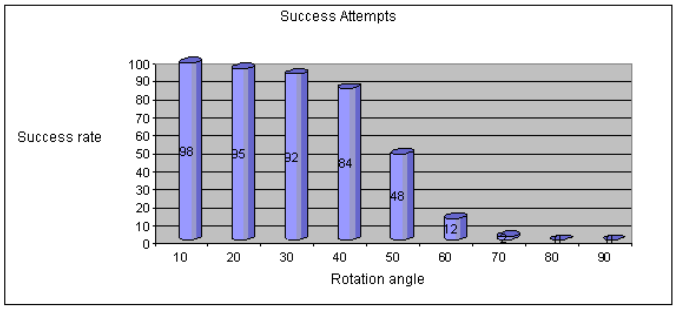

Figure 3: Robustness measured on Z-axis rotation

In the following paragraph the efficiency of the posture detection algorithm is compared in two situations: using a simple and a complex background as it can be seen from Table 1.

Table 1 Performance of the classifiers using simple and complex background images.

Clasificator	**Detected objects**	**Missed objects**	**Wrong obejct detected**
Fist using simple background.	100	0	0
Fist using complex background.	84	16	20
Palm using simple background.	100	0	0
Palm using complex background.	92	8	6

"Vision Based" approaches utilized to control a virtual hand compared to "Data Gloves" approaches. Both implemented methods control a virtual hand in real time as it can be seen from Table 2:

Table 2 Compare between vision base approach and CyberGlove-data gloves approaches

Actiune	**Method**	**Cost**	**Restriction**
Virtual hand control using a „Vision based” method	An improved method of the approach presented in „Viola and Jones”.	Reduced implementation cost. Execution cost are reduced to 0.	Hand detection will be influenced by hand postion in front of the camera.

Virtual hand control using a „Data Glove" approach	Direct metod. A driver is used to capture the data from a Cyber Glove glove.	High costs to aquire a glove.	CyberGlove sensors are very sensitives. The distance between the computer where the deviced is conected and hand is limited by cable lenght. Two hands cannot be detected on the same time, only with two gloves.

5. Conclusions

The presented method is an automatic method of hand detection and recognition first and human gestures.

The method is used to control a virtual hand with high performance and low complexity of computational viewpoint.

The method has shown good performance when it was used simple or complex background image.

6. Acknowledgements

The authors would like to acknowledge IBM Romania support and assistance for developing this research.

7. References

[1] P. Garg, N. Aggarwal, and S. Sofat. Vision Based Hand Gesture Recognition. *World Academy of Science, Engineering and Technology*. Issue 49, January 2009

[2] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen. The Columbia Grasp Database. *IEEE Int. Conf. on Robotics and Automation*. Kobe, 2009.

[3] V. Vezhnevets, V. Sazonov, and A. Andreeva. A Survey on Pixel-Based Skin Color Detection Techniques. *Proceedings of the GraphiCon*. 2003, pp. 85-92.

[4] P. Yogarajah, J. Condell, K. Curran, and P. Mc Kevitt. A Dynamic threshold approach for Skin Segmentation in Color Images. *Proceedings of 2010 IEEE 17th International Conference on Image Processing*. September 26-29, Hong Kong, 2010.

[5] Z. Hea, T. Tan, and Z. Suna. Topology modeling for Adaboost-cascade based object detection. *Journal Pattern Recognition Letters archive.* Volume 31, Issue 9, July 2010.

[6] P. Viola, and M. Jones. Robust Real-time Object Detection. *International Journal of Computer Vision*. February 2001.

[7] http://opencv.willowgarage.com/wiki.

[8] A. Miller. *Grasp it a versatile simulator*. PhD Thesis. Massachusetts Institute of Technology. 2006.

[9] http://en.wikipedia.org/wiki/Image_resolution#Spatial_resolution.

[10] http://note.sonots.com/SciSoftware/haartraining/document.html.

[11] M. Elmezain, A. Al-Hamadi, and B. Michaelis.. A Robust Method for Hand Tracking Using Mean-shift Algorithm and Kalman Filter in Stereo Color Image Sequences. *Proceedings of the International Conference on Computer Vision, Image and Signal Processing*. 2009, pp. 24-28.

[12] D. Kelly, J. McDonalda, and C. Markhama. A person independent system for recognition of hand next term postures used in sign language. *Pattern Recognition Letters*. Volume 31, Issue 11, August, 2010, pp. 1359-1368.

[13] R. Y. Wang, and J. Popovic. Real-time hand-tracking with a color glove. *Journal ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH*. Volume 28, Issue 3, August 2009.

[14] http://note.sonots.com/SciSoftware/haartraining.html.

[15] S. P. Won, W. Melek, F. Golnaraghi. A fastened bolt tracking system for a hand-held tool using an inertial measurement unit and a triaxial magnetometer. *Industrial Electronics, IECON '09, 35th Annual Conference of IEEE*. University of Waterloo, Waterloo, ON, Canada, 2009.