The 4th International Conference
"Computational Mechanics
and Virtual Engineering"
COMEC 2011
20-22 OCTOBER 2011, Brasov, Romania

REAL TIME HAND TRACKING AND GESTURE RECOGNITION FOR INTERACTIVE VIRTUAL HAND CONTROL

Catalin Constantin Moldovan[1], Ionel Staretu[2]
[1] Transilvania University of Brasov ROMANIA, mcatalin1983@yahoo.com
[2] Transilvania University of Brasov, Romania, staretu@unitbv.ro

Abstract*: Objects tracking, captured from the real word, is an area of research that has attracted a lot of attention lately, for its potential regarding the interaction between man and machines. The ability to track hand movement provides new opportunities for human computer interaction and also hand gesture recognition, in real time, and the video stream plays a significant role in controlling a virtual hand in computer generated reality applications. On the current, digital image processing applications this represent a difficult task. The present 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 optimizing gesture recognition will be to detect and track in an efficiently manner the human hand from a video stream, then detect the gesture only from a region of a picture and use the result to control a virtual hand into a virtual environment.*
Keywords: *Grasping, Motion Detection, Object Recognition*

1. INTRODUCTION

The interaction methods between man and machine had evolved over time from text based user interfaces, to 2D graphical interfaces with multimedia support, reaching today at detecting fingers from multiple persons in a complex virtual reality environment [13].
The researches were based in general on the following approaches:

- approaches based on different methods of interaction between human hand and computer using devices that digitize human hand [5]
- approaches based on different algorithms of image editing as it can be seen in [4] and [6]

A wide range of devices were developed to capture and eventually to reduce the total number of freedom degrees. These can be classified into magnetically, mechanical, optical, acoustic and inertial tracking devices. [11]
Mechanical device uses rigid or flexible goniometers, worn by user and utilized to detect hand motion. A goniometer is an instrument that either measures angles or allows an object to be rotated to a precise angular position. The goniometers from skeleton links have correspondence with the user's joints and are utilized to detect joint angle which are subsequently used in kinetic algorithms to determine the position of the body or just for a part of the body. Fixing and positioning links on the body presents some difficulties. The soft tissue of the body allows positioning the links relative with the position of the body when movement occurs. Even without these changes in the skeleton position, goniometres alignment with the body joints is a difficult task, and increases in difficulty if the number of degrees of freedom (DOF) is increased likewise, just as with human hand. Because of the variations in anthropometric measures, body-fixed systems must be calibrated for each user.
An example for the electro-mechanical device can be considered the CyberGlove system [5]. The CyberGlove motion capture system has been used in a wide variety of real-world applications, including digital prototype evaluation, virtual reality biomechanics, and animation. Cyber Glove was crated in two versions: the 18 and 22 sensor motion capture:

- The 18-sensor data glove features two bend sensors on each finger, four abduction sensors, plus sensors measuring thumb crossover, palm arch, wrist flexion, and wrist abduction.
- The 22-sensor data glove has three flexion sensors per finger, four abduction sensors, a palm-arch sensor, and sensors to measure wrist flexion and abduction. Each sensor is extremely thin and flexible being virtually undetectable in the lightweight elastic glove.

The **magnetic device**, utilized for motion capture, use sensors placed on the body for measuring low intensity magnetic fields generated by a transmitter source. The transmitter is built from three perpendicular coils, emitting a magnetic field when electrical current is applied. The current is sent to these coils in a sequence that creates three parallel fields during each measurement cycle. The detection sensors of the magnetic fields measure the power of these fields that is proportional with the distance from each coil component of the transmitter. The sensors and the source are connected to a processor that calculates the position and orientation of each sensor from the power source measuring the values of the magnetic fields. These sensors have the advantage that the measurements are correct even if the sensor is located behind a part of the body, the human body being “transparent” to magnetic fields. The main disadvantage of these devices is directly related to the physical characteristics of magnetic fields, which decrease in power if the distance from the transmitter is increased, and they can be easily altered by the presence of other magnetic fields or ferromagnetic materials.

Acoustic systems, utilized for tracking motion of the human hand, use ultrasonic pulses and determine in this way the position of the hand either by measuring the time from the transmitter to the sensor, or by triangulation. Both implementations “inside-out” and the “outside-in” are possible if this kind of devices are used, meaning that the transmitter can be positioned on the surface of the body or out of the volume that has to be measured. The sound physics restrict the accuracy of measurements and the update rate. The line formed between the transmitting device and reception sensor should not be blocked, because the measurements could be perturbed.

Inertial systems, utilize organism property to maintain the translation and rotation speed constant, excepting the case when they are disturbed by torque forces. Vestibular system, located in the inner ear is such a three-dimensional inertial biologic system. It can detect both the angular and the linear motion and it is important because it maintains the position of the body in balance by using captured data by the eye from the environment. Creation of practical inertial systems is possible with the aid of advances in micro machines and in miniaturization domains, and particularly in silicon accelerometers and rate sensors. A rate gyroscope measures the angular speed and if it is integrated, it can provide changes of the angle created with the horizontal plan, if the initial angle is known. An accelerometer measures the acceleration, including the gravitational acceleration. If the angle of the sensor formed with the vertical is known, then the gravity component can be eliminated, and by numerical integration, the speed and position could be determined. Combining the signals from inertial sensors with complementary sensors, and if the characteristics of the signals are known then the measurement errors can be minimized.

In parallel with the development of ubiquitous computing, the current methods of interaction between man and machine (keyboard, mouse, pen, data glove) are not enough anymore; the usability of these devices is limited because these are in turn limited in several ways: moving parts, contact points and so on.

Direct way to use the hand as input device is an attractive method to provide a more natural man-machine interaction.

Compared with traditional methods based on approaches that imply modifications of the device, the vision based approaches are less invasive and are more convenient to explore the three-dimensional world.

With the latest advances in the image processing, vision-based approaches become feasible in real-time interaction between man and machine.

In [7] the research is based on hand gesture tracking and recognizing and then use the data to control a virtual hand, without using a device. Researches in the area of human hand control were also made in [2], [3] and [4].

Vision based approaches must include also real time performance, accuracy and robustness requirements. Gesture detection and recognition by computer is useful in many research [10] areas like:

- Sign language recognition. Just as speech recognition can transcribe speech to text, certain types of gesture recognition software can transcribe the symbols represented through sign language into text.
- For socially assistive robotics. By using proper sensors (accelerometers and gyros) worn on the body of a user and by reading the values from those sensors, robots can assist in patient rehabilitation. The best example can be stroke rehabilitation.
- Directional indication through pointing. Pointing has a very specific purpose in our society, to reference an object or location based on its position relative to ourselves. The use of gesture recognition to determine where a person is pointing is useful for identifying the context of statements or instructions. This application is of particular interest in the field of robotics.
- Control through facial gestures. Controlling a computer through facial gestures is a useful application of gesture recognition for users who may not physically be able to use a mouse or keyboard. Eye tracking in particular may be of use for controlling cursor motion or focusing on elements of a display.
- Alternative computer interfaces. Foregoing the traditional keyboard and mouse setup to interact with a computer, strong gesture recognition could allow users to accomplish frequent or common tasks using hand or face gestures to a camera.
- Immersive game technology. Gestures can be used to control interactions within video games to try and make the game player's experience more interactive or immersive.
- Virtual controllers. For systems where the act of finding or acquiring a physical controller requires too much time, gestures can be used as an alternative control mechanism. Controlling secondary devices in a car or controlling a television set are examples of such usage.
- Affective computing. In affective computing, gesture recognition is used in the process of identifying emotional expression through computer systems.
- Remote control. Through the use of gesture recognition, "remote control with the wave of a hand" of various devices is possible. The signal must not only indicate the desired response, but also which device needs to be controlled.

In this way, the work from [8] can be considered as an example – here, the hand gestures are used as a mouse simulation mechanism. The representation is shown in Figure 1.

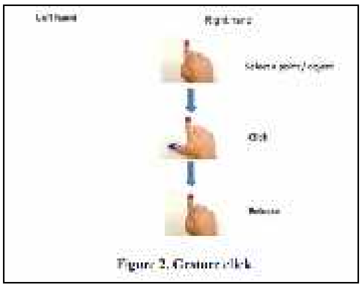

Figure 1: Mouse simulation mechanism [8]

As it can be observed, user right hand is used to simulate a mouse click. In this paper it is used an approach based on the followings steps:

- preprocessing of images (color segmentation, contours detection)
- hand gesture segmentation
- hand gesture recognition

A more efficient method is presented in [7]. The method is more efficient because it is using algorithms to detect and recognize gesture without color markers placed on hand. Human hand is detected from video stream as it can be seen in Figure 2 and used to control a virtual hand in Figure 3. Basically, the gestures are recognized based on the above schema:

- Use the real hand as an input:

Figure 2: Real hand [7]

- It can be obtained:

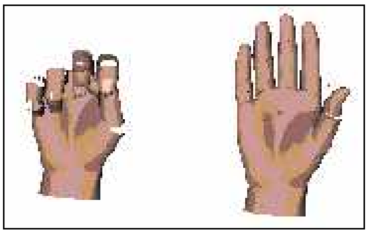

Figure 3: Virtual hand [7]

The method has satisfactory results if it is used a semi-controlled environment. For human hand detection and recognition an efficient multi-step algorithm was made:

- image segmentation between skin and non-skin region, a improved technique of [14]
- human hand posture detection, a similar approach with [1]
- human hand posture recognition
- virtual hand animation

The method proposed in [7] can be improved and extended in the following way:

- total time of execution can be reduced by eliminating the additional step of detection of skin and non-skin regions
- the control module of virtual hand can be improved by using a data glove simulation which uses the data from web camera

2. PROPOSED METHOD

The proposed method has three steps that help to control a virtual hand. These steps are:

- hand detection from an image, having as a result, tracking of human hand from a video sequence as it can be observed in Figure 4.
- hand gesture recognition represented by human hand
- use hand gesture as an input into a complex system to control a virtual hand as it can be observed in Figure 5.

In this way an application was created to track the human hand and detect various human hand gestures as follows:

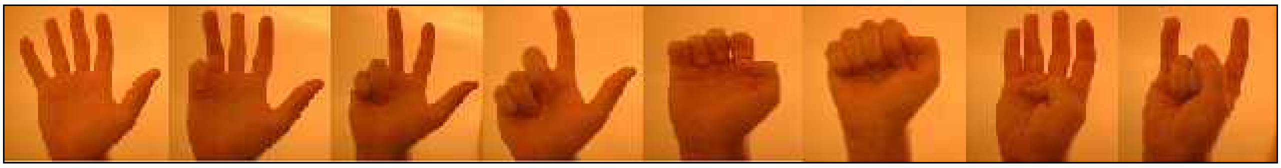

Figure 4: Human hand gestures

Using them to control a virtual hand:

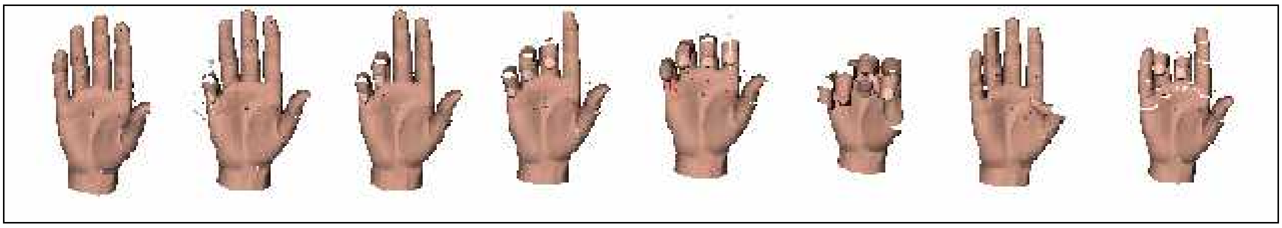

Figure 5: Virtual hand gestures

3. SYSTEM STRUCTURE

An image captured from a video stream pass into the following steps to have the hand posture detected:

- human hand detection from an image
- human hand recognition of hand posture

The following schema (Figure 6) represents the data flow, an image from a video stream pas from a video file to the package which represent the input control for virtual hand:

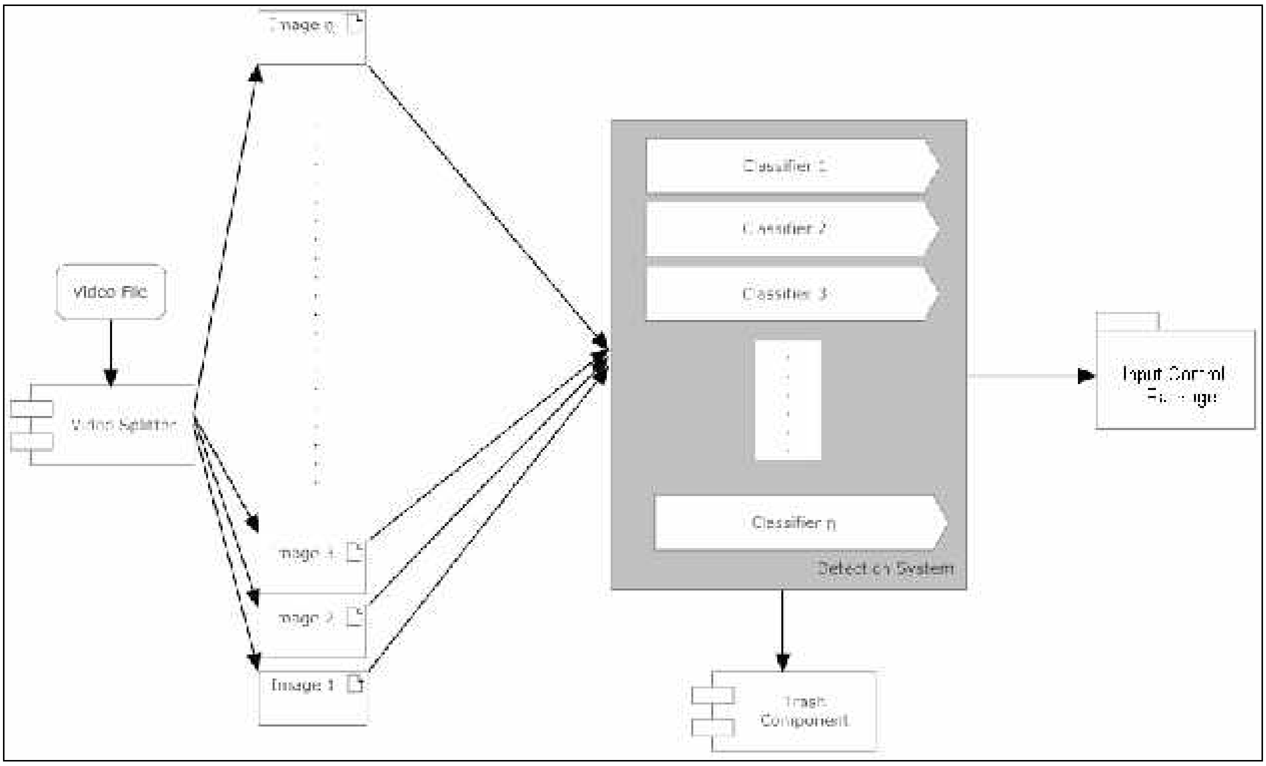

Figure 6: Hand posture recognition system architecture

The next step is represented by the schema from Figure 7. In the component displayed on the Figure 7 only positives images reach. If the algorithm does not detect any hand posture, then the virtual hand will not change its position.

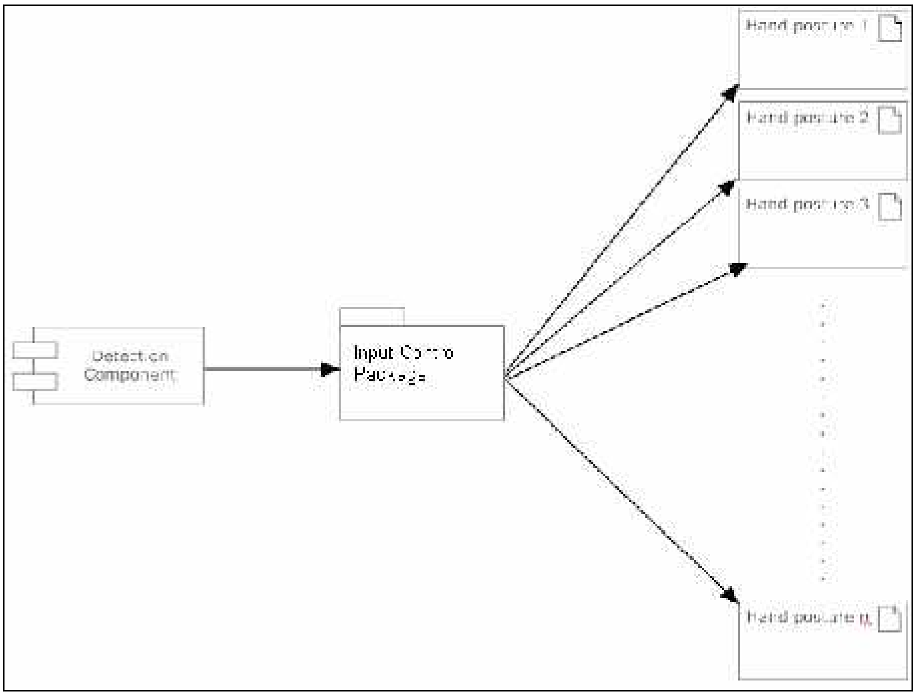

Figure 7: Control a virtual hand

4. EXPERIMENTAL RESULTS

From the experiments, it has been obtained a success rate of approximate 80% and a false rate of 20%. The false rate was influenced by couple of unpredictable factors as: inserting a new hand into the detection area, background changes, using a background similar with hand color.
In Table 1 are presented the results of the experiment:

Table 1: Results of the experiment

	Posture 1	Posture 2	Posture 3	Posture 4	Posture 5	Posture 6	Posture 7	Posture 8
Numar de incercari	100	100	100	100	100	100	100	100
Detectii reusite	95	72	80	82	23	97	91	67
Detectii false	5	22	15	3	35	0	2	15
Nicio detectie	5	6	5	15	42	3	7	8

To reduce the number of false positive, it can be applied a set of minimum requirements for frames in which a human hand will be detected.
An alternative improvement can be considered hand contour detection, or creation of a complex method of skin detection using algorithms which are not affected by changes in the background color. In this way it can be minimized the influence of the background color in the detection process.

The proposed system performs well in the both steps detection and hand simulation on virtual environment, although there is room from improvements before creating this for commercial environment.

5. CONCLUSIONS AND FUTHURE WORK

A grasping operation depends crucially on the correct view of the object which will be grasped. After obtaining the image of that object, the techniques used for image processing and analysis plays a critical role in the process of recognition the object shape and pre-configuration of the gripper.

Image processing and analysis used with the scope of gripping, can be realized using existing software technologies adapted properly to this purpose and conceive of specific software modules.
Image processing techniques can be used together with a system composed of two synchronized web camera, or by using evolved image capturing systems (Kinekt, TOF camera). In this way, a high success rate can be obtained.

5. ACKNOWLEDGEMENTS

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

6. REFERENCES

[1] Chen Q., Georganas N.D, Petriu E.: Real-time Vision-based Hand Gesture Recognition Using Haar-like Features, Instrumentation and Measurement Technology Conference – IMTC, Otawa, 2007
[2] Garg P., Naveen A., Sofat S.: Vision Based Hand Gesture Recognition, World Academy of Science, Engineering and Technology 49, 2009
[3] Goldfeder C., Ciocarlie M., Dang H., Allen P.K.: The Columbia Grasp Database, IEEE Int. Conf. on Robotics and Automation, Kobe, 2009
[4] Hea Z., Tan T., Suna Z.: Topology modeling for Adaboost-cascade based object detection, Journal Pattern Recognition Letters archive, Volume 31, Issue 9, July 2010
[5] Itu, A.: Contribuţii la strategi de prehensiune în medii reale si virtuale folosind un prehensor antropomorf cu trei degete - teză de doctorat, Universitatea Transilvania, Brasov, 2010
[6] Kelly D., McDonalda J., Markhama C.: A person independent system for recognition of hand next term postures used in sign language, Pattern Recognition Letters, Volume 31, Issue 11, August, 2010
[7] Moldovan C.C., Staretu I.: Real-time gesture recognition for controlling a virtual hand, Proceedings of 2011 International Conference on Optimization of the Robots and Manipulators, Sinaia, 2011
[8] Schwaller M., Lalanne D., Abou Khaled O.: Creation and evaluation of a Portable Gestural Interface, Proceedings: NordiCHI, October 16–20, PyGmI, 2010
[9] Staretu, I., Gripping Systems (in Romanian), Ed. Lux Libris, Brasov, 2010
[10] http://en.wikipedia.org/wiki/Gesture_recognition
[11] http://note.sonots.com/SciSoftware/haartraining/document.html
[12] Yang X.: A hand input-based approach to intuitive human computer interactions in virtual reality, Teza doctorat, The University of Hong Kong, 2010
[13] Yao J., Cooperstock J.R.: Arm Gesture Detection in a Classroom Environment, Published in: Proceeding WACV '02 Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision IEEE Computer Society, Washington, DC, 2002
[14] Yogarajah P., Condell J., Curran K., Mc Kevitt P.: 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