

The approach of the virtual reality regarding the recognition and tracking of the motion

Mihai Victor Borascu –Dumitru¹, Nicolae Popa¹

¹University of Pitesti

E-mail:npopa49@yahoo.com

Abstract. Virtual reality is a last hour technology whose main objective is to change the traditional mode of interaction with the computer in a lot of activity domains by adding new special types of peripheral devices which can facilitate the interaction with the user on all the human sensory channels.

Keywords: *virtual reality, Kinect sensor*

1.Introduction

The virtual environment is not a static one and it permanently answers to the user's commands (gestures and/or verbal commands). In the virtual environment, the user interacts with virtual objects in real time. The computer can detect the input data sent by user and can immediately modify the virtual environment based on these data. Also, the virtual reality provides the impression of presence in the 3D environment generated by the computer to the user. Another feature of virtual reality is the immersion. In this way, the user not only visualizes the virtual objects, but he also has the sensation of touching and feeling them. More than this, virtual reality technology has started to be used to a certain degree in industrial applications. So, an important purpose of the efforts which have been done in the worldwide research is to facilitate the implementation of the virtual reality in the industrial processes and to evaluate the impact and the feasibility that it has on the market and on the daily life in terms of cost effectiveness, human-machine interaction, and the side effects on the users and also the impact on the work environment at individual and organizational level. Many applications of the virtual reality presented until now have demonstrated its important potential in industry. The most advanced applications can be found in the field of space [Foust 2007] and in the field of the vehicles' industry [Tideman 2004]. The diversification of the technologies and of the applications' areas creates a real need to implement the virtual reality technology in the industrial processes. The virtual reality is a strongly interdisciplinary research domain, involving engineering (mechanical, electronics, automation, mechatronics, robotics etc.), information technologies, physics (especially optics) and the study of the human factors (cognitive psychology).

In the last years, research in field of stereoscopic visualization have strongly intensified mostly because of the exponential evolution of the computation technology but also because of the growing interest manifested by the academic community in this domain, by the professionals from various domains (like medicine, vehicle construction, architecture etc.) and manifested by the "home" users.

Thereby, in the specialty literature can be found multiple research directions, the most interesting approaches of these being presented in the following sections.

2. Virtual reality systems

- a. [Foster 2011] suggests a model for the planning and execution of the gripping based only on sensorial information in the absence of the force return and when the hand is not visible. So, the relative motion and the information related to the binocular disparity were isolated from the other depth indices and their efficiency for gripping motions and visual assessments was studied. The obtained results show that (i) the amplitude of the gripping movement grows when the relative movement is added to the information related to the binocular disparity and that (ii) both in the case of the haptic task and of the visual assessments the same distortion of the experimental derived depth was recorded. The developed model is in concordance with the results obtained by [Domini 2006].
- b. [Buckingham 2011] has done a lot of experiments with the help of which he demonstrated that the participants had difficulties in scaling the force applied with the fingertips on some objects which they had to lift without being able to see them (the objects were presented in the training phase of the experiment, but their size was different from one training session to another), demonstrating the importance of the sight as a hint for perceiving the weight of the objects.
- c. [Svarverud 2010] chose to apply rules of combining the depth hints which usually are used to perceive the shape of the surface and to issue judgments related to the location of the objects. The dipping virtual reality was used to explore the relations between different depth hints. So, that values related to changes detected in the distance where the object was located were measured in the situation when the depth hints are only “physical” (stereoscopy and motion parallax) or based on textures (independent of the size of the visualized scene). Then, they used these values to predict the distortions which appear in a task of determining the distance. The combination method of the depth hints which is used is based on totally different principles from those which are the base of the traditional models of three-dimension reconstruction.
- d. [González 2010] assumes that the conflict between different depth hints is responsible of the results showing that the convergence/divergence motion of the eyes induced by the change of disparity is not an efficient criterion for the appreciation of the depth. Estimates of the depth motion and records of the ocular motion showed that the convergence/divergence motion of the eyes induced by the change of disparity is an efficient criterion for the assessment of depth only when the rapid expansion of the image’s size is not also present. When there is a conflict between the rapid expansion of the image and disparity, then the first one represents a more powerful depth hint.
- e. [Sousa 2010] observes that when an object is stereoscopic presented in the dark, at different distances, the distance to the object is usually underestimated. Following the experiments which were performed, they concluded that adding a second object in the scene can reduce the degree in which the distance is underestimated. However, this happens only if the second object is further than that whose position needs to be estimated. So, a new depth hint is suggested, and it is the disparity relative to the furthest object.
- f. [Chen 2009a, 2009b] studies simulations of the realized sight using models of virtual reality from the point of view of the virtual reality device which was used and of the rendering mode of the phosphenes. Phosphenes are defined as any visual sensation produced by other means except for the stimulation of the visual system using the light. The results which were obtained from doing different experiments related to the functional capacity of the prosthetic sight are examined, especially regarding reading abilities, visual acuity, learning, etc.
- g. [Nawrot 2009] suggested a new formula which can link the dynamic geometry to the depth calculation based on the parallax motion. Mathematically speaking, the report between the motion of the retinal image and tracking of the eye gives the necessary information to compute the relative depth based on the parallax motion. Furthermore, it proves that the changes of the motion/tracking report are better correlated with the changes which happen in perceiving the depth based on the parallax motion than the correlation of the changes in motion or in tracking considered independently. The theoretical framework offered by the law motion/tracking gives the quantitative basis needed to study this fundamental ability of perceiving the depth.

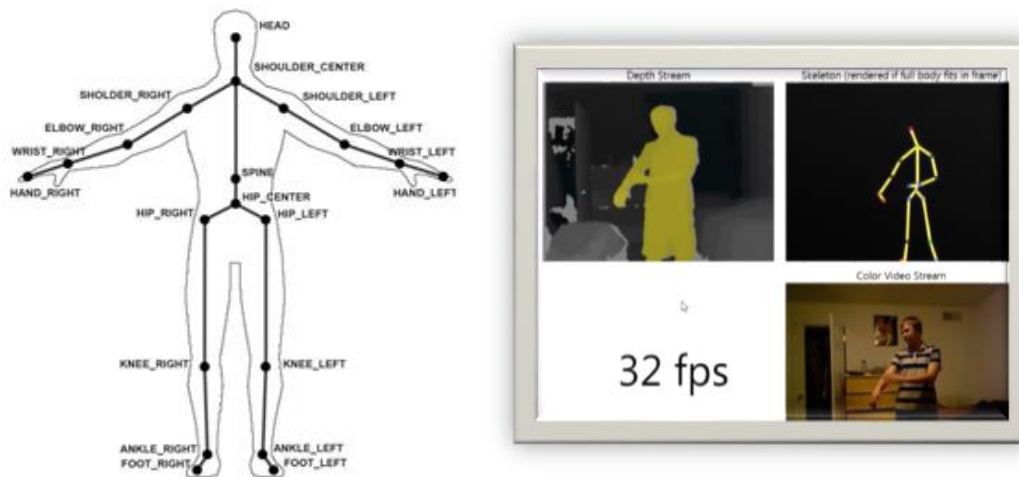
h. [Häkkinen 2006] compares the discomfort symptoms caused by simulators in three distinct situations: when the participants in the experiment use a regular screen of 17”and when HMD visualization systems are used, with the mention that in the third situation the stereoscopy was used. The results obtained in case of using the regular screen and of the virtual reality used without stereoscopy showed that there are no remarkable differences between the discomfort symptoms caused by the simulator. In stereoscopic conditions, the eye strain and the disorientation symptoms are significantly increased compared to the situation of using the regular screen.

3.Equipment for gesture recognition. Kinect sensor. Facilities and tools

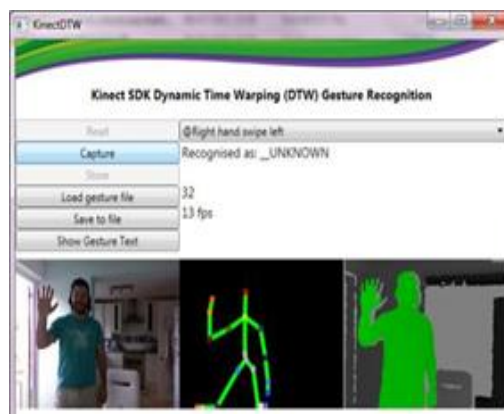
In the following section an approach based on the Kinect sensors is presented which in addition to the 2D color image also provides a depth image generated with the integrated optical sensor.

The SDK of Kinect makes available to users useful tools for developing applications having the role of tracking the head and other parts of the body and, also, for gesture recognition like: estimation of the person’s skeleton, detection and tracking of the face / head etc.

The skeleton is shaped like a list of 3D (X, Y, Z) points corresponding to 20 (Kinect-V1) wrist types („joints”) [30]: JointType.HandLeft, JointType.WristLeft, JointType.ElbowLeft etc.

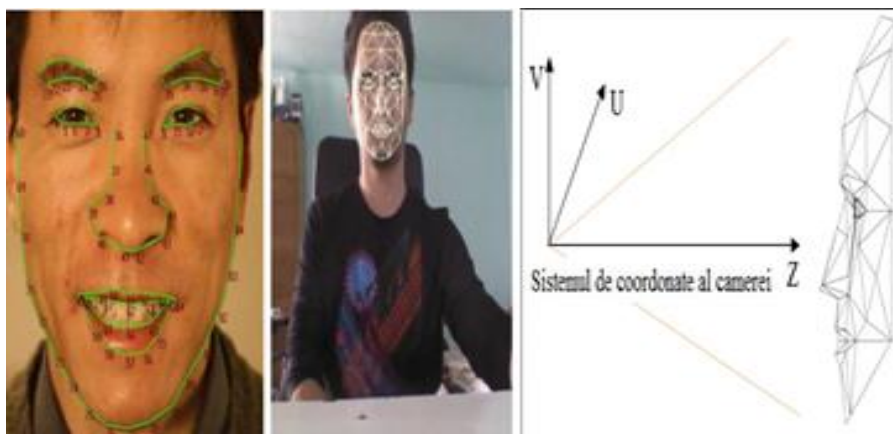


Based on the skeleton representation gestures of the tracked person can be inferred, the most used technique being DTW (Dynamic Time Warping). The implementation of a tool for the recognition of gestures related to the upper part of the body trunk by applying the DTW method on the 2D coordinates of the wrists (front-view projection of the 3D points) is described.



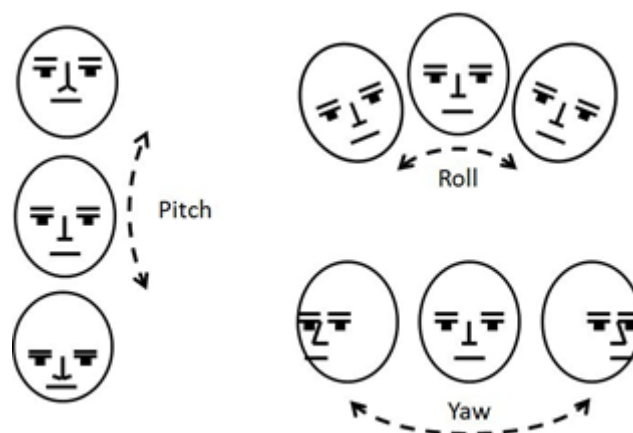
The toolkit of gesture recognition based on the skeleton and DTW [22]

Starting with the appearance of Kinect SDK 1.5, a framework was released for the detection and tracking of the face. This framework uses two libraries: Microsoft.Kinect.Toolkit and the derived library named FaceTracking. It can offer a set of 87 2D points existing on face. Additionally, the Face Tracking routine identifies other 13 points (inferred from the existing ones, for instance: the center of the eye, the center of the nose) which are not illustrated on the figure. These points can be used as primary sensorial data for different applications like HMI (Human Machine Interface), for instance: recognition of a certain facial expression, detection and tracking of hand gestures, head gestures and so on. Having the position of the joint of the head, the Kinect sensor recognizes a virtual surface (“parameterized mask” around the face based on the CANDIDE model and the Face Tracking classes return a vector of 2D points (front-view projection of 3D points) or 3D by extrapolation (according to the CANDIDE-3 model).



Points detected on the face using the FaceTracking tool (Kinect Toolkit) [21]: a. 2D points; b. 3D virtual surface obtained by extrapolation from 2D according to the CANDIDE model.

FaceTracking model provides also the *frame.Rotation* object which contains three rotation angles (*pitch, yaw and roll*) of the head around the 3 coordinate axis of the Kinect sensor and also the *frame.Translation* object which contains all the three components of the translation vector [Tx, Ty, Tz] of the head related to the same coordinate system. By analysis of these vectors' values, the position and direction where the human subject is looking can be inferred (useful in applications of monitoring the direction where the person is looking like for instance ADAS applications).



4. Conclusions

The term “**virtual**” is a concept applied in many domains and it has different meanings.

In philosophy it is related to what is not real but can provide qualities of the reality. By extension, the “virtual” term is used in technique to describe simulations made with the computer by following an existing physical model or an imaginary one. A **virtual world** shapes the real world with 3D structure and extends it with sensorial (**multimode**) perception mechanisms: visual, auditive, tactile or olfactory.

Virtual Reality (VR) as an IT technology represents a *set of technologies* which allow the interfacing of a person with an environment artificially created with the help of computer. Being concerned about a complex hardware-software system with strong psychological and cognitive implications, numerous definitions emphasize these aspects.

5. References

[1] Articles from Wiley Online Library which includes:

Tideman 2004

Foust 2007

Foster 2011

Buckingham 2011

Svarverud 2010

González 2010

Sousa 2010

Chen 2009a, 2009b

Nawrot 2009

Häkkinen 2006

[2] Alin Moldoveanu, s.a. Realitate virtuală, *Editura Matrix Rom București*, 2009.

[3] Wikipedia - Realitate virtuală, Sencorulkinect