Application of human body movements on the avatars model for the purpose of virtual training system

Branislav Sobota, Ladislav Jacho*, and Štefan Korečko

Abstract: This paper deals with particular issues in detection of human body’s movement using contactless sensor and consecutive application of recorded data on avatars model. We aimed at applying transformations to the skeleton model and skin of the avatar. Initial stages include choice of proper sensor and appropriate methods for transformation description. A semi-immersive virtual reality training system in the form of a fitting room has been chosen as a particular implementation of the body movement detection and application. The fitting room was selected because of its entertaining nature and a potential to be used as a training system for handicapped people with movement impairments.

Keywords: humanoid avatar; skeleton model; fitting room; virtual training

1 Introduction

In virtual worlds user’s identity is usually represented in the form of 3D avatar. Appropriate utilization of this phenomenon is crucial for user’s acceptance. The work proposes and implements a fundamental part of a virtual training system where the avatar is controlled with the help of human body movements. Control of skeletal system is the primary object of interest here. The goal is to create a humanoid avatar including particular transformations and the transformations are based on data from chosen sensor. In the context of this work the meaning of the word “training” can be understand in two different ways; as a training of the avatar on the basis of recorded human body movements and as using the avatar to train human users. Although the avatar is constructed to be similar to human by appearance, behavior, thinking, feeling and other features, there are big differences between human and humanoid avatar [1].

Implementation of such system includes capturing of human body movements and their mapping on 3D model of avatar. Motion capture systems found their place in several fields of our everyday life, such as entertainment, media, textile industry, medical treatment or ergonomics in automotive and many others. These systems are divided according to the requirement of markers to marker based and marker-less systems [2]. Marker based systems available from website (http://www.cs.cmu.edu/~yaser/Lecture-3-MarkerBasedMocap.pdf) use special equipment which is bound to user’s body. Traditional motion capture systems like optical, magnetic or ultrasonic require complex pre-processing and calibration in laboratory environment. The marker-less systems don’t require to wear any additional equipment during the tracking process. This is especially important in the case of handicapped users, who can be unable to manage such equipment or can be too distracted by it. There are several known techniques to obtain a skeletal model by analysis of image captured by multiple synchronized cameras [3–6] or infrared sensors [7, 8]. Authors consider price and mobility of whole solution from the start.

Our intention is to develop a system which is able to operate in real-time so the usage of data that require heavy post-processing is inappropriate. As a tracking device the Kinect sensor was chosen as one of the most affordable solutions in the field of contactless sensing of human body with respect to the price and availability and offers full support and documentation for development. On the other
hand, output of the device has to be filtered because of its relatively low and varying precision. This issue could be solved using more sensors operating through inter-process communication on one computer or through a local area network [9]. As we consider using multiple sensors in the future, we decided to utilize the first version of the Kinect sensor. The second version has considerably better features, but prohibits the use of multiple sensors on one computer.

There are many possible applications of avatars reflecting real human body movements. As one of our long-term research interests is training and education of handicapped children utilizing virtual reality technologies, we chose a virtual fitting room. This is because the activity of trying out clothes involves a natural set of movements. The system can measure variability and accuracy of these movements and evaluate abilities of a person using it. It can also guide the person to move in a particular way (e.g. to see the clothes from certain angle). In this way we can measure and train the person without his or her awareness.

The rest of the paper describes the relationship between the human skeletal system and an avatar skeletal system and its implementation in the form of the virtual fitting room. Section 2 is dedicated to existing virtual fitting room solutions which we analyzed for our implementation to be at par with the already existing ones. Section 3 introduces skeleton tracking process as implemented on the Kinect sensor. Sections 4 describes mapping of captured data to avatar’s 3D model in real time and section 5 presents underlying mathematics used for the model transformation. In addition to the choice of appropriate body tracking method it is also required to build a kinematic skeleton with skinned body model that can be animated and visualized with a possibility to apply body deformations. These issues are treated in section 6. Section 7 describes implementation details and architecture of the developed system. In sections 7 and 8 a prototype application that includes a human body movements-controlled avatar, a virtual fitting room, is presented. The paper concludes with advantages of proposed solution and outlines the way in which this system can be extended to a fully operational virtual training system.

2 Virtual fitting room

There are several applications that focus on texture mapping and creation of a virtual fitting room with a virtual mirror.

One of the best and highest quality results are provided by the 3D Body Scanner available from document (http://www.tc2.com/products/истем/), which is able to recommend the best clothes for the user on the basis of his/her body scan. The scanning can be provided by V-Dresser or Virtual Fashion. The disadvantage of the application is that it doesn’t provide a database for more users, only for the owner of the application.

Similar solution [12] and [13] is provided by the German company Human Solutions. Their body scanning system has a triangle shape base with three scanners, which uses laser to detect the body shape. As in the previous case, a disadvantage of the scanner is lack of mobility. In addition to the scanner, Human Solutions also provide an application for virtual clothes trying, called Virtual Mirror. Here after the scanning the user can choose the clothes and can see it on the screen of computer.

Products of Styku [14] pod are a virtual fitting room and a body scanner, too. The advantage of the scanning is the option to access the whole set of a scanning system, including Microsoft Kinect sensor and tablet with Windows 8 operating system. The body scanning of Styku is easily manipulated and offers steps on how to scan the body on the tablet screen with a history where the user can save all the clothes and his/her clothing size. Another application which provides virtual fitting room without scanning is Swivel available from website (http://www.facecake.com/swivel/features.htm). It is an application by FaceCake to try on not only clothes, but also accessories. It is possible to install Swivel at home too and not just use it in certain shops. It also allows taking a picture in the dress and sharing it on a social network. The only significant disadvantage is that the dresses don’t fit the user correctly and cannot follow the user movements. This is why the user doesn’t have feeling that he wears the clothes.

Clothia available from website (http://www.clothia.com) is another application for virtual fitting room, but it is only available for iPads and the user can’t see the clothes in 3D. On the other hand, it allows to take a picture and to share it on a social network. Before sharing the picture the user can move with the dress and fit it to the correct
position. However, this may be a negative aspect, since it can bother the user.

Applications like TruFit, BodyPal, MyVirtualModel, Metail and Mipso provide services for the virtual try on, too. They are available online, but the disadvantage here is that they require information from the user such as photo or clothes size. The other disadvantage is that, they are only lightly interactive.

Other applications like Me-Ality, Alvascan or Spacesurf Cartesia provide scanning option only but not the virtual mirror. What these applications have in common is that they don’t provide a model or figure type selection from database, mapping of clothes textures or objects onto skeleton and control of this model by the user motions and movements. Another disadvantage is that although there is a possibility to scan the user’s body and preview the scanned model, it is impossible to try clothes on it.

3 Skeleton Tracking with Kinect Sensor

After start of tracking, Kinect sensor provides several types of data. Human skeleton tracking is the most significant. Software of Kinect is able to determine human skeleton using cameras and IR projector. It can determine position of human standing before Kinect and also positions of particular points of human body. There are two modes of sensing: active and passive.

Active sensing provides positions of mentioned body parts shown in Fig. 1.

Positions of tracked human skeleton are provided in three-dimensional coordinates $x$, $y$, $z$ and are measured in meters. However, during active skeleton tracking it can happen that Kinect is not able to track some positions under certain conditions. The user can perform poses and gestures which Kinect is not capable to track. Examples are rotation of a hand around axis that comes through hand or head rotation to the left or to the right. In these cases Kinect could not notice a change and skeleton of the avatar could not be updated for consistency with user’s body. This is considerable limitation of chosen sensor, which causes that avatar will not be able to learn all movements that human can do.

4 Humanoid Skeleton Representation

Avatar’s skeletal system has to be pre-defined. In computer graphics, mainly in the field of 3D character animation, skeletal system is represented as mathematical tree. In graph theory, tree is an undirected graph where two nodes are connected by an edge. In the case of avatar’s skeleton nodes represent joints (fixed or movable) and edges represent bones between them.

When we apply geometric transformation on particular joint we are considering this joint and its belonging bone as one unit. Center of the rotation is placed in the joint where a bone is connected. There are hierarchical relationships in a tree, which means that a transformation applied on one unit could affect previous units as well.

Transformations are represented by matrices. In considered skeletal model each of units has a relative matrix assigned, which specifies relative position to the parent unit and an absolute matrix specifying absolute position of the joint in a virtual scene. Relative matrix of particular unit is multiplied by absolute matrix of the parent and the result is the absolute matrix of this unit.

Hierarchical structure of skeleton system in tree graph was described using XML.

4.1 Starting Poses

Detection of starting poses is used to start updates of avatar in virtual scene according to tracked user’s body.
As soon as virtual training system detects defined pose, computation of absolute matrices for all units of skeletal system is performed and avatar is brought to ready state.

4.2 Conversion of sensor data

Microsoft Kinect provides positions of certain parts of human body. It is necessary to calculate rotation of these parts to control avatar in appropriate way. At first vectors from pairs of positions are calculated and then their comparison is performed to provide rotation of particular part. Each of these vectors represents a bone of human body and can be assigned to corresponding bone of avatar. It is calculated as subtraction of two positions of body parts.

When two positions are available $P_1$ with coordinates $x_1$, $y_1$, $z_1$ and $P_2$ with coordinates $x_2$, $y_2$, $z_2$, then vector $V$ can be defined as subtraction of positions $P_1 - P_2$. There are three components of $V$, $x = x_1 - x_2$, $y = y_1 - y_2$ and $z = z_1 - z_2$. This vector represents direction and length of particular bone of human skeleton. Corresponding avatar’s bone has to be transformed from starting pose to such a pose, that vector created from endpoints of this bone will be identical with vector of human bone.

To achieve this goal, the transformation of vectors will be represented as quaternions. In mathematics, quaternions are used to extend the system of complex numbers and they are not commutative. They found application in physics, applied mathematics and also in computer graphics. For purposes of this work, quaternion is described in terms of computer graphics and used to describe rotary transformations.

5 Application of Transformations on Avatar Model

Quaternions include two items; one is a vector part that describes an axis of rotation and scalar part that describes amount of rotation. To get required pose of avatar’s bone, it’s necessary to rotate this bone by defined angle around an axis.

To define this quaternion, it is necessary to determine rotation axis and rotation size.

Axis of rotation can be determined by cross product of avatar’s bone vector and human bone vector. Prior to calculating cross product is necessary to normalize both vectors. Cross product is binary operation of two vectors and result is also vector, which is perpendicular to both source vectors or to plane, which is defined by these vectors. Output is described by Eq. 1.

$$ O = V_1 \times V_2 = \begin{vmatrix} I & J & K \\ v_{10} & v_{11} & v_{12} \\ v_{20} & v_{21} & v_{22} \end{vmatrix} \tag{1}$$

This can be further expressed in Eq. 2.

$$ \begin{vmatrix} I & J & K \\ v_{10} & v_{11} & v_{12} \\ v_{20} & v_{21} & v_{22} \end{vmatrix} = (v_{11}v_{22} - v_{12}v_{21})I + (v_{12}v_{20} - v_{10}v_{22})J + (v_{10}v_{21} - v_{11}v_{20})K \tag{2}$$

Axis of rotation is $O$, $V_1$ and $V_2$ are vectors of bones and $I$, $J$ and $K$ are base vectors of Cartesian coordinate space.

After determination of rotation axis, it is necessary to determine rotation size, angle. The angle between two vectors can be calculated using their dot product. Dot product is directly proportional to cosine of angle between these vectors which is further defined in Eq. 3.

$$ \theta = \arccos(V_1 \cdot V_2) \tag{3}$$

$\theta$ represents required angle and $V_1$ and $V_2$ are normalized vectors of human bone and avatar’s bone. Rotation axis $O$ is described in Eq. 4.

$$ O = [x_0, y_0, z_0] \tag{4}$$

Considering this equation and rotation angle marked as $\theta$, quaternion $q$ can be defined as in Eq. 5.

$$ q = \left( x_0 \sin \left( \frac{1}{2} \theta \right), y_0 \sin \left( \frac{1}{2} \theta \right), z_0 \sin \left( \frac{1}{2} \theta \right), \cos \left( \frac{1}{2} \theta \right) \right) \tag{5}$$

After application of every bone rotation described by this quaternion we can get an avatar to identical pose as the pose performed by the user. Transformation of avatars bone is shown in Fig. 2.

6 Skinning and Deformation

Visualization of results includes skin deformations of avatars model. Preparation of avatar model was realized using Autodesk 3ds Max. Each vertex of avatars mesh can be associated with bones which indicate the strength of deformation. All bones contain inner and outer envelopes to affect vertexes using weight. The weight linearly decreases from 1 to 0 according to direction of movement.
Skin deformations are also affected by smoothing modifiers to improve lack of credibility raised during computation process because of intersection of particular polygons. Results of skin deformation are shown in Fig. 3. Envelopes of a bone are represented by red outlines. Red color also presents parts which are affected by higher weight. Places where mixing of weights occurs are joints of bones, which are yellow. There are also black places which usually represent unsatisfying deformations.

7 Implementation

The application of virtual fitting room was developed to evaluate results. It represents Windows based application with the possibility to try various clothes on a 3D avatar responsive to user’s moves. Proposed system was developed using Microsoft .NET Framework version 4.0 and Microsoft Visual Studio 2013 as integrated development environment. The application uses two software libraries including Kinect SDK which supports methods for extraction of human body from depth map and XNA Framework for implantation of 3D scene. There are alternatives, to XNA such as Monogame and SlimDX, however they don't support loading models in form of compiled files. There are several core parts of XNA framework, which was used for implementation of the key features:


Application design including data exchange is shown Fig. 4.

Usage of graphical data in compiled form during runtime eliminates latency required to determine the asset file format and extract data from this file into a format that application can process.

As it was stated in the beginning, virtual training system is the main goal of this work. Secondary purpose of developed application (virtual fitting room) was chosen to extend attractiveness and enjoyment in the use of the whole system. Fulfillment of requirements for such application includes the ability to manage significant graphical data in the form of 3D models with meshes and materials as well as 2D thumbnails representing clothes. There are several solutions for storage of these graphical elements.

One of the possibilities is the usage of spatial databases or also geodatabases (GDB), which are aimed at this kind of data. Several database management systems including MySQL support GDB as installable extensions. Stored data in the form of linked table entries are easy to access because of known nature of relational databases.

Chosen software technologies provide possibility to use XNA framework to access graphical models and textures using its interface. XNA framework includes so-called Content Pipeline (Fig. 5.), a set of functions and processes to import assets from a source files and handle them through their attributes and methods.

We consider having rich graphical content in the future so combination of both approaches seems to be advantageous. Appropriate solution should contain determination of models which will be loaded from GDB and subsequently accessed by the user. Models and textures should be stored in compiled form. This solution can be optimization of initialization of the whole system as well as decrease memory requirements on the start. For the pro-
Application of human body movements on the avatar

Figure 4: Virtual fitting room application design.

As it was mentioned in section 4 and 5, preparation of avatars model includes definition of its skeleton and additional parameters. To be able to work with such graphical data in developed system it is necessary to import and process them in appropriate way. 3D models are imported per a standard Autodesk FBX importer, offered by XNA.

We minded to develop application software performing animations of models, so the usage of build-in Model Processor was insufficient as it is aimed at model processing only for rendering. Therefore a new processor named Skinning Processor was implemented by extending of the Model Class.

Skinning processor is responsible for validation of individual nodes of the object model. The first step is determination of the node. There are two types of nodes, a bone or a mesh. In the case that a node is determined as a mesh it is necessary to check geometry attributes required for skin deformation process. Validation process is unsuccessful in the case the attribute describing weight value absent. After the validation the Skinning Processor searches for a node representing the root bone of the avatar’s skeleton. This bone is used as a reference point for animation of the whole model. This part is also used to compute transformation matrixes of the model as it was proposed in section 4.

8 Application usage

Application of fitting room is supposed to accomplish two core functionalities from the user’s view. Avatar’s model selection is the first. There is a space for improvements in the meaning of addition of other models which will express the users figure by his opinion the most. Another use case is a choice of desired clothes from available collection. In this case cloth material is mapped on avatar’s model, which is responsive to user’s moves. Full list of use cases is described by the diagram in Fig. 5.

The user navigates through collection using hand gestures (swipe left and swipe right). Graphical user interface includes buttons in the overlay layer. Interaction starts if the user holds a hand over them for a while, as it is known from applications with Kinect support. This action is interpreted to the system as a click. System is supposed to be controllable mainly by body movements but it also supports keyboard input. Keyboard navigation is used to ma-
nipulate with camera position in each direction and to zoom in and zoom out the view. It is possible to turn on or turn off avatar’s mirroring and also switch to sitting mode (i.e. the avatar sits).

Usage of developed application is shown in Fig. 7.

9 Conclusion

Virtual training system prototype presented here is in its initial stages based on the virtual fitting room. Our work consists from achieving two main goals. One of them is tracking human body movements using contactless sensor. Proper sensor choice and tracking technique was necessary to fulfil this part. Second goal consists of preparation of captured data to be applied on human avatar model including expression of skin deformations. We also paid attention to a case of a virtual fitting room to extend purpose of our developed system.

Used method brought appropriate results; however there are inaccuracies in motion tracking which could be improved using more than one sensor Kinect or with the use of RGB camera which is embedded in this device.

Our future works includes extension of the virtual fitting room to a virtual training system that trains a human user via a virtual avatar. Prove of concept includes its deployment in educational process of handicapped people. As the clothes and fashion represent an attractive area for children, especially for girls, the fitting room system can be adjusted for a therapy and evaluation of motoric skills of handicapped children. In this way the children could see the usage of developed system as a kind of entertainment and not an education (training). Whole solution will be used for capturing movements, evaluation and proposing of corrections for the user in the form of animated avatar.

Acknowledgement: This work was supported by the KEGA grant no. 083TUKE-4/2015 “Virtual-reality technologies in the process of handicapped persons education”.

Figure 6: Use case diagram of the virtual fitting room application.

Figure 7: Virtual fitting room application.
References


