A Virtual Object Manipulation System Using Human Body Postures

Yonemoto, Satoshi
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

https://hdl.handle.net/2324/5805

出版情報：Proceedings of International Conference on Applied Computing, 2004-03
バージョン：
権利関係：

A VIRTUAL OBJECT MANIPULATION SYSTEM USING HUMAN BODY POSTURES

Satoshi Yonemoto
Faculty of Information Science, Kyushu Sangyo University
2-3-1 Matsukadai Fukuoka 8138503 Japan

Rin-ichiro Taniguchi
Department of Intelligent Systems, Kyushu University
6-1 Kasuga-koen Fukuoka 8168580 Japan

ABSTRACT

This paper describes a virtual reality system that realizes both an avatar motion control and a virtual object manipulation without avatar representation. Our goal is to do seamless mapping of human motion in the real world into virtual environments. We hope that the idea of direct human motion sensing will be used on future intelligent user interfaces. Our method can generate realistic avatar motion from the sensing data. We use virtual scene context as a priori knowledge. We assume that virtual environments provide action information for the avatar. In the virtual object manipulation without avatar representation, the user behaviors through physical-virtual interaction are interpreted as the object manipulation tasks. The behaviors which the user performs in the real world are converted into the scene events to manipulate the interesting virtual object.

KEYWORDS

image processing, motion capture, perceptual user interfaces

1. INTRODUCTION

Use of 3-D human motion sensing without physical restrictions is the most promising approach to realize seamless coupling between virtual environments and real world. In particular, the goal of our research is to do seamless mapping of human motion in the real world into the virtual environments. Motion capturing by computer vision techniques is applicable for such purposes (Yonemoto,S.,2000, Wren,C.,1997). We have already developed the vision techniques to capture the user motions without any specific markers.

As a first step, we have developed an avatar motion control by user body postures. It is very beneficial to estimate 3D body postures. We think that realistic presentation of the substitute body is important subject to percept self body in the virtual scene or to communicate with the other existence. Our method is based on motion synthesis from a limited number of perceptual cues, which can be stably estimated by vision process. In the virtual environments, basic postures of the avatar can be represented by these methods. In addition, we use virtual scene context as a priori knowledge. We assume that virtual objects in the virtual environment can afford avatar's action by simulating the idea of affordance in the real world. In other words, the virtual environments provide action information for the avatar, such as properties of the virtual objects. An important point here is that we can consider scene constraints in the virtual scene to generate more realistic motion beyond the limitation of the real world sensing. Every task in the manipulation is strongly related to objects in the virtual environments.

As the second step, we have developed a virtual object manipulation interface without avatar representation. In this case, the user can handle hand cursors in the virtual environment. The user motions of the head and both hands are captured by the vision process which we have already developed. The user behaviors through physical-virtual interaction are interpreted as the object manipulation tasks. In other words, the behaviors which the user performs in the real world are converted into the scene events to manipulate the interesting virtual object.

2. HUMAN BODY MOTION TRACKING

In order to extract a rough sketch of human body from the image, or blobs are mainly used as effective visual features. We also employ skin color regions of a face and hands in the image. Figure 1 shows an example of 2D blob tracking results. When a 2D blob is detected in two views, or in multiple views, the 3D position of the blob can be calculated by stereo vision. The algorithm of 3D blob position calculation is summarized as follows: According to camera calibration, for each of the views, a line of sight, or a vector from the origin of the camera coordinate system to the center of mass of the blob, is calculated. Intersecting the lines of sight, the 3D position of each blob is calculated. Grasping gestures are also estimated observing the change of hand region.

3. VIRTUAL OBJECT MANIPULATION

Our system supports both realistic avatar motion control by user body postures and virtual object manipulation without the avatar representation. The former case is necessary in direct virtual object manipulation by the user or in the situation that the other users watch the user actions through networks (the internet). "The direct" means that the user motions should be realized in the virtual environment as close as possible. The latter case is employed when changing the states of the virtual object is the unique goal for the user. In this case, the user motion does not have to be represented realistically.

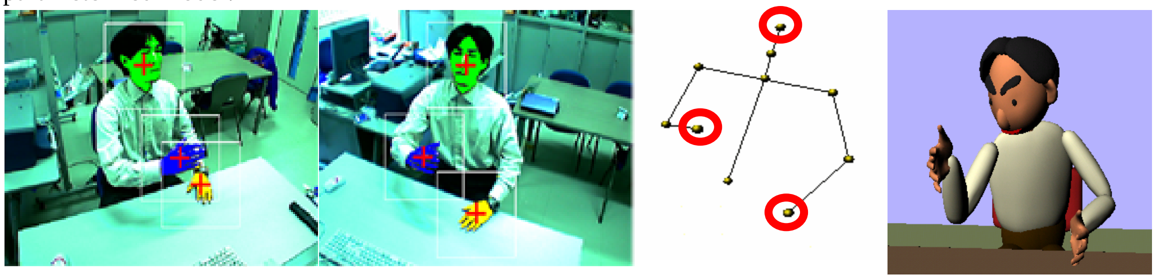
3.1 Avatar Motion Control by User Body Postures

Though many motion synthesis methods are proposed, we need to develop a method which can control an avatar from the sensing data. We have introduced a physically constrained motion synthesis, where only the skeletal model of the avatar obeys the physics laws which satisfy tight point-to-point constraints between adjacent joints. Here, we assume that a link between adjacent joint points is represented by a strong spring model so as to keep the link length constant. Moreover, we assume that the velocities of the joint points can be approximately calculated from the displacements of the joint positions. All joint positions that were estimated by iteration process keep the balance according to the change of the positions of the head and the hands. The other constraints can be also added to the iteration process. Figure 2 shows the reconstructed skeletal model.

We assume that each virtual object affords essential information about user's action based on simulating the idea of affordance in the virtual environments. In other words, we assume that interaction among the virtual objects and the user should be properly performed by making each virtual object give rise to afforded action. Both the states of the virtual objects and the states of the avatar are decided whether the virtual objects are handled or not, according to the distance between the virtual objects and the user. For example, when the user grasps a virtual cup, the afforded finger motions to grasp it are augmented and the state of the object is changed *static* into *move*. Such augmented motions are not acquired by real-world sensing but automatically provided from the virtual objects. In addition, our framework can consider scene constraints in the virtual environments to simulate scene events realistically. Since the virtual scene is completely recognized by the system, the context reasoning in the interaction through the virtual environments is not more serious than in the real world sensing.

3.2 Object Manipulation without Avatar Representation

The user can handle the hand cursors in the virtual environment. Only the hand cursors are displayed for the user. The user motions of the head and both hands are captured by the vision process. The user behaviors through physical-virtual interaction are interpreted as the object manipulation tasks. In other words, the behaviors which the user performs in the real world are converted into the scene events to manipulate the interesting virtual object. For example, in the case which the user handles a door fixed in the virtual scene, the user's motion is converted into a force to move the door. The motion of the door is represented as a parameterized model.



(a) **(b)** **(c)** **(d)**

Figure 1. Motion sensing. (a)(b)An example of 2D blob tracking results from two viewpoints (c) a reconstructed 3D skeletal model using three blob positions (d) avatar representation

4. EXPERIMENTAL RESULTS

We have applied our methods to desktop-style system, which is implemented on one PC. The system installs two cameras and a wide 2D display in front of the user. Figure 2 shows the system setup. The system can perform real-time from vision process to virtual scene rendering. The user can only monitor the projected virtual scene with the 2D display.

4.1 Avatar Motion Control

The user can handle the virtual objects with the intension of achieving his tasks. Our system supports the following user tasks in the virtual environments.

(a) ***transfer object to a target position***

In this case, the user task is decomposed into the sequence of sub-tasks: "reach towards the handle position", "grasp the handle", "transfer towards the target position", and "release the handle". Based on our avatar motion control, augmented motions such as the face direction control and the finger motions are automatically added into the basic avatar motions.

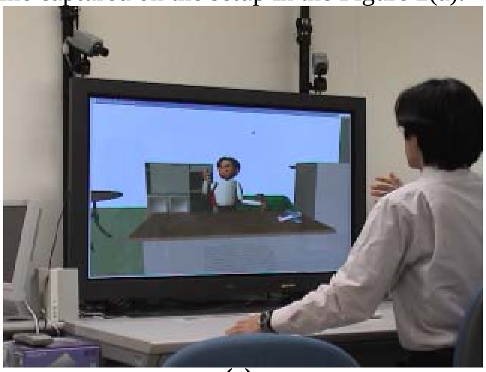
(b) ***change the states of object, fixed in the virtual environment***

In this case, the user task is decomposed into the sequence of sub-tasks: "reach towards the handle position", "grasp the handle", "change the states of object", and "release the handle". Augmented motions such as the face direction control and the finger motions are added. In changing the states, for example, opening/closing the door, the constrained motions are forced into circular along the knob position.

(c) ***use the functions of object***

In this case, the user task is decomposed into the sequence of sub-tasks: "reach towards the handle position", "grasp the handle", "transfer towards the target position", and "release the handle". The target means the other object to which applies the functions of the object. Based on the functions of the object, augmented motions such as the face direction control, finger motions and the wrist motions are added.

Many manipulation tasks are the typical cases of "transfer object to a target position". Manipulation tasks such as reaching, grasping and transferring are needed. As perceptual cues, 3D positions of head and both hands, and appearances of hands related to grasping motion are acquired. Figure 2(b)(c) shows the object manipulation tasks. First, the user performs reaching tasks. According to the afforded actions, he or she manipulates the interesting virtual objects. The afforded actions involve the additional motions of manipulation tasks. For example, in grasping, finger motions which cannot be estimated from a limit number of perceptual cues are supplemented to augment the virtual scene. The face direction of the avatar is automatically controlled. In the example of Figure 2(b), the avatar tracks the grasping object. The shots are online captured on the setup in the Figure 2(a).



(a)

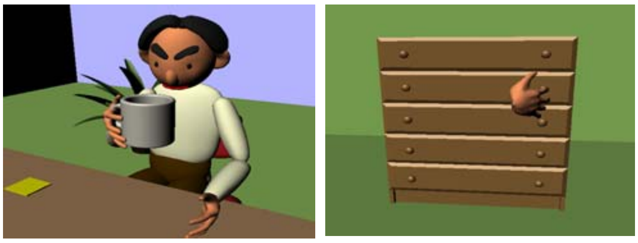


Figure 2. The object manipulation system.
(a) The system setup
(b) Manipulation with avatar control
(c) Manipulation without avatar representation

4.2 Object Manipulation without Avatar Representation

We tested handling a door and a drawer fixed in the virtual scene. Such objects are represented as parameterized models. Grasping the handle (i.e. the knob), the avatar hand is fixed in the virtual scene. While the user grasps the knob, the user's motions converted into a force to move the object. When the user grasps them, the relative motions can be easily exchanged into the rotation of the object in the virtual environments.
The viewpoints are freely controlled using the head motion or the hand motion. The different point of the related researches is that the user motions do not have to correspond with the cursor position in the virtual scene. Since our approach uses the vision data, the system needs to inference the user's intention from the rough motion sequences.

5. CONCLUSION

We realized both an avatar motion control and a virtual object manipulation without avatar representation. Our framework is based on unwired sensing, so the user can easily experience seamless coupling between the real world and the virtual environments. Avatar motion control is augmented, simulating the idea of affordance in the virtual environments. In the virtual object manipulation without avatar representation, the user behaviors through physical-virtual interaction are interpreted as the object manipulation tasks. Future works include personal adaptation with the avatar and the application with more complex interaction tasks.

ACKNOWLEDGEMENT

This work has been partly supported by the programs of the Grant-in-Aid for Young Scientists (B) from the Japan Society for the Promotion of Science (15700172).

REFERENCES

Douville, B. et al, 1996, Task-level object grasping for simulated agents. *Presence: Teleoperators and Virtual Environments*, Vol.5,No.4, pp.416-430.

Kuffer,J.,1999, Autonomous Agents for Real-time Animation, PhD thesis Stanford University.

Koga, Y., 1994, Planning Motions with Intentions, *Proceedings of SIGGRAPH'94, pp.24-29.*

Wren, C. et al, 1997, Pfinder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 19,No.7, pp.780-785.

Yonemoto,S. et al, 2000, Real-time Human Motion analysis and IK-based Human figure Control. *Proceedings of Workshop on Human Motion*, pp.149-154.