Using Motion Capture for Real-time Augmented Reality Scenes
Hiromitsu Sato & Michael Cohen
Spatial Media Group
University of Aizu
Aizu-Wakamatsu, Fukushima-ken 965-8580
Japan
{m5131106, mcohen}@u-aizu.ac.jp

ABSTRACT
The purpose of this study is improvement of real-time human-computer interface in using augmented reality. We propose animating an avatar in real-time with motion capture system and augmented reality.

Development of ubiquitous computing is impressive in recent years. Augmented reality is an instance of ubiquitous computing technology. Augmented reality is related to virtual reality, characterized by composition of real and virtual, and display of electronic information as virtual object and giving the real world addition of information.

Because computer performance is continually upgrading, we can display complicated 3D models, and it is possible to easily add information to a computer such as a mobile telephone. Core technologies of augmented reality are position recognition, image recognition, and orientation recognition. In the past, high performance machines were needed to process these. Using modern hardware, we can process such data using mobile phone as well as notebook. Recently, Sekai Camera became famous as augmented reality software that runs on iPhone and Android. Assuming continual computer performance upgrade, there will appear a lot of augmented reality software we can use in the future. However, the means to affect information in augmented reality have not fully developed yet.

It is reasonable to suppose that we can use motion capture as technique to let a real human affect a figurative 3D object in augmented reality. Motion capture typically involves tracking parts of the body by infra-red cameras to obtain streamed position data. We can realize real human movement by applying this data to a 3D model. Development of devices and technology to capture motion advances. Reasonable devices and software have become widely used. For example, Kinect on the input device for Xbox 360, sold since 2010 by Microsoft, allows manipulation of a 3D model on the computer without a controller. In the future, techniques to handle 3D objects will grow more popular.

The merit of both augmented reality and motion capture is to enable dynamic manipulation a virtual space. It is likely that one will be able to exchange one’s own information for another’s in asymptotically realistic environments.

An avatar was produced which works on augmented reality in real-time as proof of concept. When it was developed, efficiency was emphasized by using motion capture software “Vicon IQ,” motion capture real-time engine “Tarsus,” C language library “ARToolKit,” and 3DCG software “3ds max.”

Categories and Subject Descriptors
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems– Artificial, augmented, and virtual realities.
I.4.8 [Image Processing and Computer Vision]: Scene Analysis – Tracking.

General Terms
Measurement, Design, Human Factors

Keywords
3DCG, augmented reality, motion capture, OpenGL

1. RELATED TECHNIQUES
For example, Microsoft Kinect (for the Xbox 360 video game platform by Microsoft), Wii wireless remote controller (the primary controller for Nintendo's Wii console), and Playstation Move (the motion-sensing game controller for the PlayStation 3 video game platform by Sony Computer Entertainment) are approaches for producing avatars animated with augmented reality.

However, these systems have problems regarding range of movement and recognition accuracy. When users move in small motions, sometimes they cannot be recognized. Additionally, because range of capture is not so wide, user cannot move broadly.

For addressing these problems, marker tracking systems use additional cameras and markers. One of the common methods of recording movements is motion capture.

2. METHOD
2.1 Merit using Motion Capture
It is difficult to express realistic human movement in 3DCG.
Because various elements such as muscle, human frame, and the placement of center of gravity have influence on a human motion, it is very complex. That’s why to create this movement to use 3DCG software precisely is difficult. Models animated manually often make an unnatural impression.

Motion capture is a technique for digitally recording movements. Motion capture is called “mocap” for short.

In the area of entertainment, we record motion to realize human motion. In the context of filmmaking, “mocap” refers to the technique of recording the actions of human actors with the intention of using that information to animate digital character models in 3D animation.

There are various kinds of motion capture, with merits and demerits associated with each. We outline what is used commonly.

If the movement of a model lacks realism, it disappoints as avatar performance. Therefore, a technique to realize realistic human movement expression is necessary to perform avatar movement with some information.

### 2.1.1 Mechanical
For mechanical sensing, the actor wears rigid parts made of metal on the body. Mechanical motion capture is invoked, as this exoskeleton moves and a sensor recognizes the movement when its actor moves.

The first merit is there is no interference from magnetic or light fields. Also, the detection of movement is certain.

A demerit is that a mechanical sensing system does not know which way a performer's body is pointing unless some other type of sensor in place.

### 2.1.2 Optical
For optical sensing, a performer wears retro-reflective dots that are tracked by multiple cameras emitting infra-red rays. Each marker is wrapped in reflective material.

There are a number of merits to this type of motion capture. First merit is performer feels free to move because he need not connect his body to equipment or cables. Next, it is possible to use larger spaces, and record more performers’ movements. Finally, we can get very clean, detailed data using special software. In this research, we use Vicon IQ.

There are some demerits. First, reflective markers can be blocked by performers or other structures, causing loss of data. But the occlusion can be compensated for with software which estimates the position of missing points. Next, performer must wear a suit with markers and balls, which may be uncomfortable. Finally, an optical motion capture system is more expensive than other types of motion capture.

The UBIC (University-Business Innovation Center) motion capture room at the University of Aizu holds in total 12 optical cameras, which can surround a model. Because there are many cameras, the movement of a whole body can be captured without limits. Therefore, optical motion capture system was used in this research.

### 2.1.3 Electromagnetic (magnetic)
A magnetic type is system that generates a magnetic field in the measurement range, and measures the change of the magnetic field where it is affected by the movement of the measurement object. It is necessary that an actor wear an array of magnetic receivers which track position.

The merit is that there is no blind spot when performer exercises, and it is relatively low cost.

First demerit is that a magnetic field is distorted, and that influences accuracy when metal is near the capture area (Cement floors usually contain metal.) Second demerit is that the performer cannot move freely because he needs to wear cables connected to a computer.

### 2.2 Motion Capture Data File Format
The BVH format is motion capture data file format which Biovision Company originally developed. The name BVH stands for “Biovision hierarchical data.” Some commercial 3D software such as MotionBuilder (real-time 3D character animation software developed by Autodesk Media and Entertainment), 3ds max (modeling, animation, and rendering package developed by Autodesk Media and Entertainment), and Poser (3D CGI rendering and animation software developed by Smith Micro Software) support the BVH format now.

The main BVH format features are the following:

- The BVH format is described ASCII.
- The coordinate system is right-handed.
- The handling of each XYZ axis is left to the user.
- The BVH format describes information about the joint nodes.
- The joint turns are described in Euler angle form.
- The unit of rotary angle is degree.
- The HIERARCHY part describes the skeleton hierarchical structure of the character.
- The MOTION part describes movement data.
- It is impossible to write comments.

Because the coordinate system of the BVH format is right-handed, the BVH format and OpenGL fit together well. In contrast, the coordinate system of DirectX is left-handed, so there is a need to convert coordinate transformations.
Figure 1: BVH format example

Beside, the CSM format is an ASCII file used to import positional marker data from various motion capture systems into Character Studio to animate bipedal characters. We can export CSM file to describe the movement of a whole body by measuring 30 (up to 49 with an option) measurement points of the human body (see Table 1).

Table 1. Character Studio-Supported Marker Names

<table>
<thead>
<tr>
<th>Body Part</th>
<th>Code</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>LFHD</td>
<td>Left Front Head</td>
</tr>
<tr>
<td></td>
<td>LBHD</td>
<td>Left Back Head</td>
</tr>
<tr>
<td></td>
<td>RBHD</td>
<td>Right Back Head</td>
</tr>
<tr>
<td></td>
<td>RFHD</td>
<td>Right Front Head</td>
</tr>
<tr>
<td>Chest</td>
<td>CLAV</td>
<td>Top Chest</td>
</tr>
<tr>
<td></td>
<td>STRN</td>
<td>Center Chest</td>
</tr>
<tr>
<td>Waist</td>
<td>LFWT</td>
<td>Left Front Waist</td>
</tr>
<tr>
<td></td>
<td>LBWT</td>
<td>Left Back Waist</td>
</tr>
<tr>
<td></td>
<td>RBWT</td>
<td>Right Back Waist</td>
</tr>
<tr>
<td></td>
<td>RFWT</td>
<td>Right Front Waist</td>
</tr>
<tr>
<td>Spine</td>
<td>C7</td>
<td>Top of Spine</td>
</tr>
<tr>
<td></td>
<td>T10</td>
<td>Middle of Back</td>
</tr>
<tr>
<td>Left Leg</td>
<td>LKNE</td>
<td>Left Outer Knee</td>
</tr>
<tr>
<td></td>
<td>LANK</td>
<td>Left Outer Ankle</td>
</tr>
<tr>
<td></td>
<td>LMT5</td>
<td>Left Outer Metatarsal</td>
</tr>
<tr>
<td></td>
<td>LTOE</td>
<td>Left Toe</td>
</tr>
<tr>
<td>Right Leg</td>
<td>RKNE</td>
<td>Right Outer Knee</td>
</tr>
<tr>
<td></td>
<td>RANK</td>
<td>Right Outer Ankle</td>
</tr>
<tr>
<td></td>
<td>RMT5</td>
<td>Right Outer Metatarsal</td>
</tr>
<tr>
<td></td>
<td>RTOE</td>
<td>Right Toe</td>
</tr>
</tbody>
</table>

2.3 ARToolKit

Augmented reality (AR) is a field of computer research which deals with the combination of real-world and computer-generated data (virtual reality), and the overlay of virtual computer graphics imagery on the real world (see Figure 2). In industrial and academic research areas, AR has many potential applications.

ARToolKit is a C language software library for the creation of augmented reality applications. It uses computer vision techniques in order to calculate the real camera position and orientation relative to physical markers in real-time, allowing one to over-lay virtual objects. It was developed at the University of Washington HIT Lab. The ARToolKit can be freely downloaded from its official web site. The source code is offered with an open source license.

The main functions that ARToolKit supports are the following:

- acquisition of images from cameras,
- detection of markers and recognition of patterns,
- measurement of three-dimensional locations and orientation of the markers,
- display compositing three-dimensional images and real images.
3. DISCUSSION
We asked some volunteers to fill out a questionnaire adapted from the experiments. Their suggestions for improvement are the following:

- Difference between web camera and observer's eye causes a feeling of strangeness.
- They have limited moving range.
- It's boring to use only fixed web camera.
- Calibration of camera requires great care.
- The program provides user with no feedback.
- The display is too small to manipulate.

We use web camera in this research. It is necessary to try not to exaggerate a feeling of strangeness attributed to difference between web camera and observer's eye. A head-mounted display would allow a more natural perspective.

4. CONCLUSIONS AND FUTURE WORK
Because we adopted an optical motion capture system to address motion range limitation, it is available to have a big range and improve the accuracy of the program. However, one problem is that calibration camera and wearing retro-reflective dots require great care. To solve these problems, it was necessary to change from optical motion capture to simple one. It is possible to maintain high-quality accuracy and broad motion range by using multiple cameras in ARToolKit.

We use ARToolKit to enable augmented reality in this research. In the future, we need to change to a marker-less augmented reality and we can move web camera freely.

We can solve the problem that the program provides someone with no feedback by using wireless controllers, and building a vibrating system into controllers. When a 3D object is touched, controllers can vibrate at the same time to provide immediate feedback. When one drags the pointer over the information one wants to select, it is available to use a controller. We don’t have to raise our hand each time we choose from options.

Because it is necessary to use in real-time, display is influenced by the performance of computer and camera. The display tended to fail to render when we used weak machine. However, recent machine specifications are much higher than the machine on which we experimented, and reproduction is possible with modern notebook PCs. When complex 3D model shapes are displayed, it is necessary to reduce the number of polygons and the size of texture maps.

In this research, real-time avatar using augmented reality was produced (see Figure 4). To increase functions, it is more useful.

5. ACKNOWLEDGMENTS
We thank all members of Special Media Group, and all my friends, for advice, cooperation, and encouragement in this research.

6. REFERENCES


