

Immersive Performance Training Tools Using Motion Capture Technology

Jacky Chan

Department of Computer
Science

City University of Hong Kong
83 Tat Chee Avenue,
Kowloon, Hong Kong
+852-21942925

jcpchan@cityu.edu.hk

Howard Leung

Department of Computer
Science

City University of Hong Kong
83 Tat Chee Avenue,
Kowloon, Hong Kong
+852-27887234

howard@cityu.edu.hk

Kai Tai Tang

Department of Computer
Science

City University of Hong Kong
83 Tat Chee Avenue,
Kowloon, Hong Kong
+852-21942837

itjeff@cityu.edu.hk

Taku Komura

School of Informatics

The University of Edinburgh
IPAB JCMB, Kings Buildings,
Mayfield Rd, Edinburgh
EH9 3JZ, U.K.
+44 (0)131 651 3445

tkomura@inf.ed.ac.uk

ABSTRACT

Traditionally, students can undergo performance training by imitating teachers' motions and improving their moves by following teachers' advice. However, teachers are not always available and there is a lack of effective self-training tools. In this paper, we propose an immersive performance training system with motion capture technology. Our aim is to provide a virtual training environment for performance training. In the environment, the student's performance is analyzed by our system and a virtual teacher will give feedback for his/her improvement. As a result, the student can self-train and get feedback the same way as in a traditional performance training lesson. We will present our prototype implementation of a dance education system according to our design of immersive performance training system with motion capture technology.

Categories and Subject Descriptors

J.5 [Performing Arts], K.3.1 [Computer Uses in Education]

General Terms

Design, Human Factors, Measurement, Performance

Keywords

Immersive VR application, Human Computer Interaction, 3D Human Motion Analysis, Performance Training.

1. INTRODUCTION

With the recent advance in computer graphic and motion capture technology, immersive virtual reality (VR) applications gain much attention in both academics and industry. Virtual environment is becoming more and more realistic. Through an immersive VR application, people may experience a real-world-like or an imaginative virtual world. Different kinds of immersive VR applications have been developed for entertainment, martial arts, etc [1][2].

An immersive training system is potentially a complement to real-world training. It can simulate an environment that may be unreachable or unrealizable in the real world. Trainee/learner can gain some experiences before they actually start training or learning in the real world, especially for trainings that may be dangerous or impossible to carry out in the real world such as military training. US army had invented an immersive VR

application that their soldiers could learn performing checkpoint duty [3]. Chou *et al.* applied virtual reality in civil engineering education that provides an alternative for students to learn structural analysis as these concepts are difficult to understand when they are visualized by 2D wire-frame diagrams [4].

Some immersive VR applications target on motion training. With motion capture technology, user can interact with the computer by body movement. Chua *et al.* proposed an immersive VR motion training system using motion capture technology for practicing Tai Chi [5]. The avatar of the Tai Chi learner and his/her instructor are put together with in a virtual environment. The student learns by mimicking the instructor's motion and knows what is wrong with their motion by observing his/her own movement from the avatar. Hachimura *et al.* proposed a dance training system with mixed reality [6] that merge motion capture and virtual reality techniques. The head mounted display shows the expert's dance motion in front of a trainee like a "Ghost". The learner can observe the 3D animation demonstrated by the expert and move his/her body until perfectly overlapped with the expert's 3D avatar. In other words, a learner should follow the postures of the expert. In these two applications, although learners can observe their own movement by observing the virtual avatar, they may not have enough experience to identify whether they have performed wrongly. In our design, learners can concentrate on their training while our system will be an observer and provide feedback to them.

The paper is organized as follows. The proposed performance training system is described in section 2. A showcase of our dance education system is presented as an example performance training application in section 3. Conclusions and future work are provided in section 4.

2. PROPOSED SYSTEM DESCRIPTION

The objective of our system is to provide a virtual environment for performance training. In this paper, we aim to teach people to dance properly. We use a motion capture system to capture the 3D motion of the student. Meanwhile, an animated avatar is rendered in a virtual classroom. There is also a virtual teacher demonstrates the moves to the student and gives feedback after analyzing the student's performance. The student is immersed in the virtual environment and feels like a real training lesson.

Figure 1 shows how a user interacts with our proposed system. Our system consists of three parts: 1) Training Room, 2) Virtual

Classroom and 3) Data Server. Training room is the area where training really takes place. It contains all necessary equipments such as motion capture system, speakers, and screen, etc. The equipments provide a sense of immersion to the student and by the way collect his/her motion data. Virtual classroom is a virtual environment where displays the virtual teacher and the student's animated avatar. It renders both pre-stored and real-time animations. Data server stores the data of the performance lesson includes the motion templates, performance names, and performance types, etc.

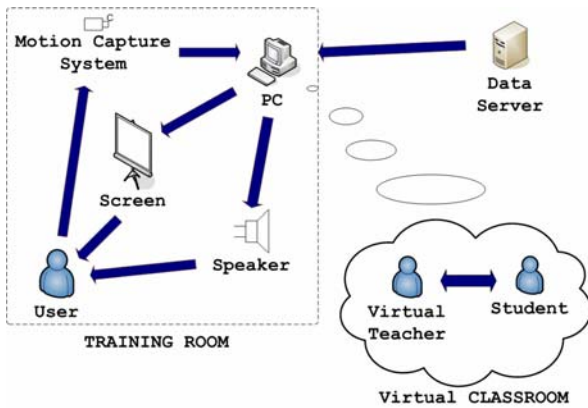


Figure 1. The user interaction diagram of our proposed performance training tool

2.1 Training Room

Figure 2 shows the detailed user interaction in the training room environment. In a training room, various equipments support the interaction between the student and the virtual teacher. During the lesson, the student's motions are captured as input data, and hence rendered by an animated avatar. Meanwhile, the student could observe the teacher's moves through the screen and listen to the instruction through the speakers. By the way, the student receives feedback.

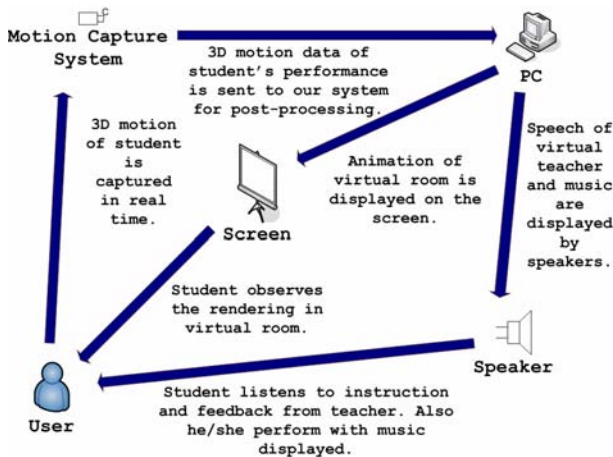


Figure 2. The user interaction diagram in the classroom environment

2.1.1 Real Time Motion Capture

In our system, we use optical motion capture system. Figure 3 shows our helper who wore a suit with markers attached and

surrounded by motion capture cameras. Those cameras are mounted on surrounding walls of the training room. The student has to wear a suit attached with reflective markers. When the motion capture system is running, the camera will emit light and track the positions of the markers through the reflected light. Combining different views from all cameras, the 3D coordinates of markers are obtained. When the training starts, the system captures the 3D motion data of the student in real time manner and the data is continuously sent to our system for processing.



Figure 3. An optical motion capture system with cameras and a suit with markers attached

2.1.2 Display

The system renders the virtual classroom and displays on the screen. The virtual teacher demonstrates the standard moves and provides feedback to the user about how good he/she had performed. Moreover, the student's animated avatar is rendered according to the captured motion data in real-time manner. As the student can control his/her own avatar in the virtual classroom, the student will feel him/herself immersed into the virtual environment.

2.1.3 Sound

Speakers are setup in the training room as some performances involve music. During each practice or demonstration session, the corresponding music can be played in the training room.

2.2 Virtual Classroom

Virtual classroom is a place where the virtual teacher and students interact. In the virtual classroom, the virtual teacher will react according to the student's action and the student will follow the suggested training plan.

2.2.1 Setup

3D motions of real teachers were captured and pre-stored in the server to form template lessons. At the beginning, the student can see the name and type of available performance lessons and choose one of them to perform training. After selecting the lesson, the corresponding data will be retrieved from the data server. At the same time, the system will check whether the motion capture system is online and ready to capture the student's motion. When everything is ready, the virtual classroom will be established.

2.2.2 Interaction

Interaction between the student and the virtual teacher is one of the important features that makes the student immersed into our system and helps him/her to achieve better performance. The

virtual teacher will give suggestions about the student’s training based on his/her current performance. The virtual teacher may either suggest a training plan or specify particular moves for the student to practice.

2.2.3 Personal Training Plan

The teacher in a real performance lesson will suggest a personal training plan for students. Students will practice basic moves first. After they get familiar with these basic moves, the teacher will teach them to combine the learned moves. Students may demonstrate different talents when performing various moves. For example, some students may have difficulty in moving legs quickly so they should focus more on practicing the leg motions.

The trainings in our system are designed in a sense that follows the normal flow in a real performance training lesson. There are three stages of trainings: 1) motion primitives training, 2) intermediate training and 3) whole motion sequence training. After analyzing the student’s motion, our system can identify the weaknesses of the student and suggest him/her to train weak motions more often.

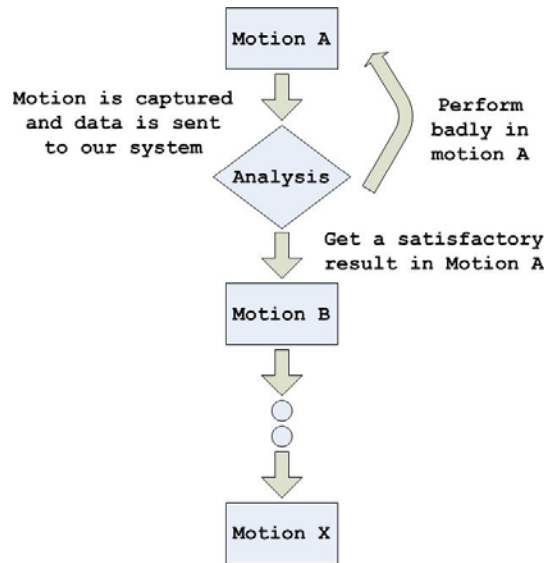


Figure 4. Flowchart of Stage 1: motion primitives training

2.2.3.1 Stage 1: Motion Primitives Training

In the first stage of each lesson, the virtual teacher will demonstrate the basic moves of a performance. The basic moves, or *motion primitives*, are ordered according to the difficulty level that depends on the difficulty of body movement. Figure 4 shows the flow of motion primitives training. After the student had completed learning one move, the virtual teacher will demonstrate next move and the student should follow the way the teacher performed. If the student does not perform well, the virtual teacher will tell the student how to improve until he/she can master the move. This stage will continue until the student can do well in all motion primitives.

2.2.3.2 Stage 2: Intermediate Training

In the second stage, the student is required to practice longer motion segments. These motion segments are parts of the whole motion sequence and they are formed by combining motion

primitives. By performing training on motion segments, students will be able to link several motion primitives together. If the student fails to perform particular motion segment, our system will point out the erroneous motion segment and break it down into smaller sub-segments that aids the student to practice. If the student still performs the sub-segment motion poorly, the system then prompts the student to go back to the first stage for motion primitives training again.

2.2.3.3 Stage 3: Whole Motion Sequence Training

In the last stage, the student attempts to perform the whole motion sequence. Figure 5 shows the flowchart. The student may perform the motion incorrectly in several ways such as forgetting some moves, mixing the order of primitive motions or having difficulty in linking up the motion segments. The system will identify and locate the error when the student performs the whole motion sequence. The virtual teacher will suggest the student to go back to the previous stage and practice the motion segments that contain significant errors. On the other hand, if there are no significant erroneous motion segments, then the student will receive an evaluation about his/her overall performance. The student can keep practicing in this stage to improve the smoothness of his/her movement. The training is complete when the system satisfies with the student’s performance.

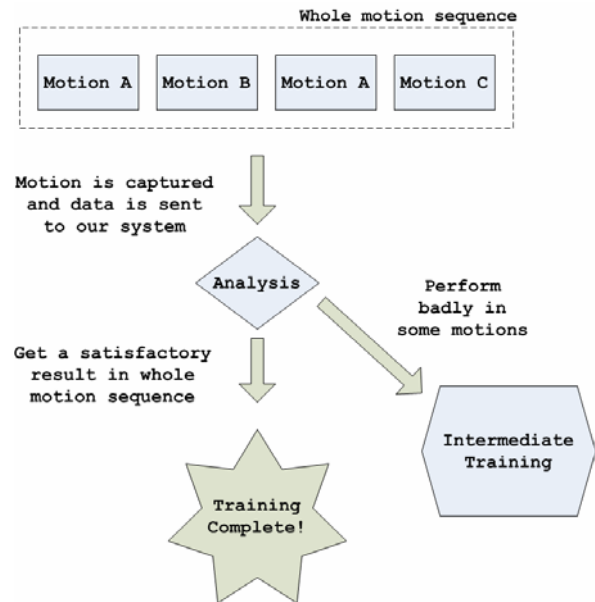


Figure 5. Flowchart of Stage 3: whole motion sequence training

2.2.4 Feedback

Throughout the training, motions of student are captured and evaluated. The system will identify erroneous joints and provide suggestion in the training plan. The feedback is provided in the form of text dialogue. This will give students the impression that the virtual teacher is human-like. The text feedback like “Your movement on LEFT HAND is not correct” is more descriptive and direct than a simple similarity score. From this kind of feedback, the student can realize their errors immediately and know how to correct the errors. Figure 6 shows the feedback from virtual teacher on the dialog box.



Figure 6. An example of feedback dialogue by virtual teacher

2.3 Data Server

Two types of data are stored in the server for defining a lesson: 1) lesson data and 2) 3D avatar models. Before the training, our system retrieves the required data from the server.

2.3.1 Lesson

Each template lesson contains motion data, name, performance type and background music. The lesson data is pre-stored. Motion data is prepared through capturing the motions performed by professionals. The motion data will be analyzed and divided into motion segments and motion primitives that are important moves and repetitive in the whole motion sequence. Motion data, motion primitives and motion segments will be stored as template motion.

2.3.2 Avatar Model

Different kinds of outfit are suitable for different types of performance. To make our virtual teachers look more professional and include more varieties, the system offers different kinds of avatars to suit different performances. For example, a clown avatar would be suitable for performing funny motion as shown in Figure 6.

3. APPLICATION SHOWCASE: DANCE EDUCATION SYSTEM

We have implemented a dance education system as a prototype of the proposed immersive performance training system. We choose dance education for our application because dance is one of the most popular performances and whole body is involved in the movement. The virtual teacher demonstrates the standard way of dancing (the template motion) and then the student is instructed to perform likewise. The 3D motion of student is captured by the optical motion capture system. Figure 7 shows a student learning dance through our dance education system. The screen shows a virtual teacher on the left hand side and the student's own virtual avatar on the right hand side. The student is trying to imitate the teacher's motion. The system keeps capturing and updating the student's poses that drives the avatar to move accordingly. The motion data will be analyzed and the virtual teacher will feedback the student what is wrong with his/her performance.

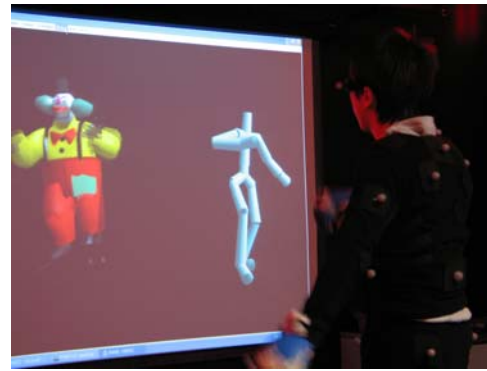


Figure 7. An example showing a student undergoes training with our Dance Education System

3.1 Rendering

The 3D animation is rendered by OpenGL, which is a popular 3D graphic toolbox that supports most platforms like C++. The rendering result is projected onto the screen.

3.2 Demonstration of Moves

Template lessons are pre-set by the real teacher. The lesson data contains the lesson name, description, dance style and template dance motion performed by the teacher. In a dance motion sequence, some moves are often repeated and we mark them as motion primitives. Template dance motions are divided into different motion primitives. The dance motion data are sorted by difficulty that determines the frequency and range of movement. Figure 8 shows a 3D virtual teacher demonstrating certain moves.



Figure 8. Dance demonstration by a 3D virtual teacher

3.3 Analysis and Feedback

Nakamura *et. al* proposed a dance training tools with a timing-vibro device and a robotic screen [7]. These two devices are used for showing the timing information and the amount of translation in a dance performance. Students' motions are animated in real-time on the screen and they get feedback on their movement pattern. This application is similar to our proposed design in the sense that we also provide real-time animation and feedback for the captured motion.

Hachimura *et. al* proposed to use *weight*, *space* and *time* for analysis and evaluation of dancing movement [8]. Weight relates to the kinetic energy of each part of the body. Space refers to the direction of movement of body as a whole. Time incorporates the rate of movement, i.e., the acceleration of each part of the body. By extracting these features from the two motions to be compared,

the time instants at which the two motions differ can be found. This approach performs a global matching between the postures but does not localize the mismatches.

In recent researches, angles between joints or different parts of the body are used for analyzing specific dance motions. For the Japanese dance Furi, four spatial indices are calculated for quantifying the characteristic features [9]. These four indices show significant differences between an expert's motion and a beginner's motion. A gesture recognition engine in a dance system also presents similar design [10]. It matches two gestures by considering the joint angles of 10 body parts including head, torso, upper arms, forearms, upper legs and lower legs but ignoring hands and feet. The similarity between two gestures is calculated based on the Mahalanobis distance. Kwon *et. al.* designed a motion training system using visual and body sensors [1]. In their approach, the body sensor is attached to the forearm for measuring two angular values, pitch and roll, in a motion. They compare similarity of two motions by using the Euclidean distance metric. Particular joint angles are considered in above researches for analysis of specific dance motions. We would like to consider more joint angles of the whole body and develop a matching algorithm for handling more general dance motions.

In our approach, the motion capture system is used to collect the motion data whose joint angles are compared in a frame-by-frame analysis. A frame contains information about a set of joints specifying a posture. The motion of the virtual teacher, or the template motion, is displayed on the screen whereas the motion of the student is captured in real time. Since the student is trying to imitate the virtual teacher's motion, each frame of the student motion in terms of joint angles should be compared with the corresponding frame of the virtual teacher's motion being displayed. After analysis, feedback will be given by the virtual teacher to the student. Different follow-up actions may be triggered depending on the analysis result at each stage.

3.3.1 Comparison of Joint Angle

Each frame consisted of joint angles of the student's motion is compared with the corresponding frame of the virtual teacher's. A frame is indeed a representation for a posture. Posture matching is performed for each pair of joint angle sets contained in the frames. Many joints are considered in the matching process. For example, torso is taken into account since the rotation along the body's principle axis is important in dance performance such as ballet. For comparing the template motion and the student's captured motion, in total 33 pairs of joint angles are used. The relative change R_{ij} for the i -th joint angle between the template motion and the student's motion at frame j is computed by equation (1):

$$R_{ij} = |B_{ij} - A_{ij}| / A_{ij} \quad (1)$$

where B_{ij} is the i -th joint angle of the j -th frame of the student's motion, A_{ij} is the i -th joint angle of j -th frame of the template motion, and $i = 1, 2, 3, \dots, 33$.

A validity index matrix \mathbf{V} is computed by the set of R values and a satisfactory index S . The elements of the validity index matrix V_{ij} for the i -th joint at frame j is computed by equation (2):

$$V_{ij} = \begin{cases} 1 & R_{ij} < S \\ 0 & R_{ij} > S \end{cases} \quad (2)$$

The satisfactory index S is set to be 0.2. The validity is set to be 1 if the relative change is smaller than the satisfactory index meaning that the move of the particular body part at the particular time instant is valid. In other words, the satisfactory index specifies the error tolerance when the matching is not perfect. There are two reasons for the imperfect matching results. First, even when two people try to perform the same posture, there will be some differences due to their unequal body shapes. Second, there may be some delay fluctuations when a student imitates the virtual teacher's moves shown on the screen. The value of 0 for V_{ij} indicates that the i -th joint at frame j is poorly matched.

3.3.2 Follow-up Actions at Each Stage

In the first stage of training, the student practices on each motion primitive. There are two possible follow-up actions based on the analysis result. The first kind of actions is to give advice on improving the particular motion primitive. After the analysis, we can identify the validity V_{ij} that is equal to 0 meaning that the move is invalid. The subscript i specifies which joint moves wrongly so the erroneous limb can be identified. The second kind of actions is to demonstrate the next motion primitive. When a motion primitives are marked as correct, i.e., when all elements of the validity matrix \mathbf{V} are equal to 1, the student is considered to be familiar with that motion primitive. The teacher will thus demonstrate a new move or start the second stage.

In the second stage, the student is asked to perform a motion segment by linking several motion primitives. Following a similar process as in the first stage, we locate those frames that contain errors by checking which element in the validity matrix V_{ij} is equal to 0. The student will be notified about his/her errors and be asked to perform the same motion segment again. If the student demonstrates recurring errors while performing motion primitive X , then the follow-up action is to bring the student back to the first stage and focus on training for the motion primitive X .

In the third stage, the student is required to perform the whole motion sequence. Similar to the previous two stages, we first locate those frames that contain errors by checking for the values of 0 in the validity matrix \mathbf{V} . Then we identify the motion segments that contain those frames. In this case, the follow-up action is to ask the student to go back to the second stage and work on the identified motion segments. On the other hand, if no errors are reflected by the validity matrix, then the student can get an evaluation about his/her overall performance. The follow-up action in this case is to let the student practice the whole motion sequence again to improve the smoothness. The training is complete when the student is satisfied with the performance.

4. Experiment

To test the usability of our prototype system, we have invited 9 subjects to try out our system and fill in the questionnaires afterward. Some of the questions are listed below at Table 1. Subjects rated for each question in order to show if they agree/disagree the proposition. For the graphs shown in Figure 9, the x -axis corresponds to the user score and the y -axis corresponds to the number of users giving the score. It can be seen in Figure 9

that all the question got score 4 or above showing that all the subjects are satisfied with our system.

Table 1. Some example questions in the survey

No.	Questions
3	The system can help you know what and where mistakes are in your dancing.
5	The virtual teacher animations are interesting and they can stimulate your interest to use this system and to observe their motions.
6	You prefer learning dancing with the system to traditional "watching dance video" approach.
9	The interaction part in the system assisted you in dancing training process.

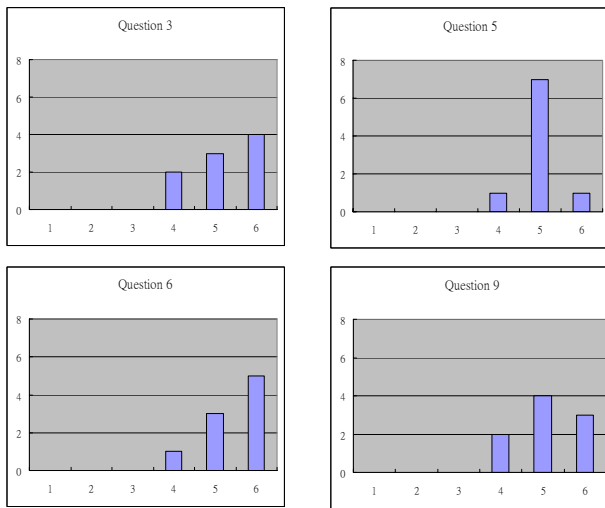


Figure 9. Subjects' response for survey's questions

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed an immersive performance training tool using the motion capture system. It provides a virtual environment for students to undergo performance training. With the motion capture system, a student can control his/her animated avatar in the virtual classroom and interact with the virtual teacher who will assist him/her in the training process. Based on the analysis of the captured 3D motion of the student, the virtual teacher will give feedback for the student to make improvement. The virtual teacher will point out mistakes made in the student's move, identify the student's weakness and suggest him/her to pay attention on particular limbs for particular moves.

As future work, the template lesson generation will be automated by considering pattern recognition techniques. Moreover, we will explore various ways to improve the matching algorithm. We will take into account the fluctuations in delays when the student is imitating the virtual teacher's motion. The analysis result will be further exploited to provide even more meaningful feedback. For example, when the student performs a wrong move at the knee, our system will be able to identify if his/her leg is required to bend more or less.

6. ACKNOWLEDGMENTS

The work described in this paper was fully supported by a grant from the Research Grants Councils of the Hong Kong Special Administration Region, China (Project No. CityU 1167/05E).

7. REFERENCES

- [1] Ronald Sidharta and Carolina Cruz-Neira, "Cyclone Uppercut, A Boxing Game for an Immersive Environment", *Proc. of the 2005 ACM SIGCHI Intl. Conf. on Advances in computer entertainment technology*, pp.363-364, 2005.
- [2] Zhenyu Yang, Bin Yu, Ross Diankov, Wanmin Wu and Ruzena Bajscy, "Collaborative Dancing in Tele-immersive Environment", *Proceedings of the 14th annual ACM intl. Conf. on Multimedia*, pp.732-726, 2006.
- [3] R.Bowen Loftin, Mark W.Scerbo, Frederic D. MchKenzie and Jean M. Catanzaro, "Training in Peacekeeping Operations Using Virtual Environments", *IEEE Computer Graphic and Applications*, Vol. 24, pp.18-21, July 2004.
- [4] Chien Chou, Hsieh-Lung Hsu and Yu-Seng Yao, "Construction of a Virtual Reality Learning Environment for Teaching Structural Analysis", *Computers in Industry*, Vol. 51, pp.31-40, May 2003.
- [5] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins and Randy Pausch, "Tai Chi: Training for Physical Tasks in Virtual Environments", *Virtual Reality, 2003. Proc. IEEE*, pp.87-94, 2003.
- [6] Koaburo Hachimura, Hiromu Kato and Hideyuki Tamura, "A Prototype Dance Training Support System with Motion Capture and Mixed Reality Technologies", *Proc. of the 2004 IEEE Intl. Workshop on Robot and Human Interactive Communication*, pp.217-222, 2004.
- [7] Akio Nakamura, Sou Tabata, Tomoya Ueda, Shinichiro Kiyofuji and Yoshinori Kuno, "Dance Training System with Active Vibro-Devices and a Mobile Image Display", *2005 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, 2005 (IROS 2005)*, pp.3075-3080, Aug 2005.
- [8] Kozahuro Hachimura, Katsumi Takashina and Mitsu Yoshimura, "Analysis and Evaluation of Dancing Movement Based on LMA", *2005 IEEE Intl. Workshop on Robots and Human Interactive Communication*, pp.294-299, 2005.
- [9] Mitsu Yoshimura, Norio Mine, Tamiko Kai and Lsao Yoshimura, "Quantification of Characteristic Features of Japanese Dance for Individuality Recognition", *Proceedings. 10th IEEE Intl. Workshop on Robot and Human Interactive Communication*, pp.193-199, Sept. 2001.
- [10] Gang Qian, Feng Guo, Todd Ingalls, Loren Olson, Jody James and Thanasis Rikakis, "A Gesture-Driven Multimodal Interactive Dance System", *2004 IEEE Intl. Conf. on Multimedia and Expo (ICME 2004)*, pp.1579-1582, June 2004.
- [11] Doo Young Kwon and Markus Cross, "Combining Body Sensors and Visual Sensors for Motion Training", *Proc. of the 2005 ACM SIGCHI Intl. Conf. on Advances in computer entertainment technology*, pp.94-101, 2005.