A Dance Training System that Maps Self-Images onto an Instruction Video

Minoru Fujimoto, Masahiko Tsukamoto
Graduate School of Engineering
Kobe University
Kobe, Japan
minoru@stu.kobe-u.ac.jp, tuka@kobe-u.ac.jp

Tsutomu Terada
Graduate School of Engineering
Kobe University
PRESTO, JST
Kobe, Japan
tsutomu@eedept.kobe-u.ac.jp

***Abstract*—Owing to recent advancements in motion capture technologies, physical exercise systems that use human interaction technologies have been attracting a great deal of attention. There are already various approaches in place that support motion training methods by using motion capture technology and wearable sensors to analyze body motion. In this study, our basic idea was to change the appearance of a dancer in an instruction video to that of the user, who we assume would be interested in seeing what they would look like if they could perform so well. We developed a motion training system that maps the user's image onto an instruction video. Evaluation results demonstrated that our proposed method is effective for motion training in specific situations.**

***Keywords*-motion capture; motion training**

I. Introduction

Owing to recent advancements in motion capture technologies [1], human interaction technologies specialized to physical exercise [2] have attracted a great deal of attention. Motion capture technologies are popular these days, and can commonly be seen in games in which users practice dance motions. Ideally, such training systems have both a high level of motion training functionality and a high entertainment value.

There are various methods in place for analyzing body motions in motion training systems [3][4], such as image processing, motion capturing using markers, and calculating body movement by using wearable sensors. For example, Kwon et al. proposed a motion training system that uses a wireless acceleration sensor and image processing [5]. Trainers and trainees can analyze their movements by watching a hybrid representation of visual and wearable sensor data that is automatically generated as a motion training video.

In our study, the basic idea is to change the appearance of an expert dancer in an instruction video to that of the user. Users, usually, practice dance movements by following the teacher's movements: they virtually translate each of the teacher's body parts into their own, and consider how best to perform the movement. In contrast, with our system, we assume that players would be interested to see their dance moves as if they could already perform really well. We propose a new practice style in which a composite video is created that shows a user dancing as if he or she were an expert. This enables users to experience what it feels like to dance skillfully. Being impressed by the performance of experts is a significant motivator for learning new techniques, so we decided to utilize this motivation for learning and support in the form of showing beginners what they would look like if they were an expert.

In the proposed motion training system, images of the user are mapped onto an instruction video. The user then practices the dance movements by following their own expertly dancing image on the video. Evaluation results demonstrated that the proposed method is effective for motion training in specific situations.

The remainder of this paper is organized as follows. In the next section we discuss related works, and in Section 3 we describe the proposed method. In Section 4, we evaluate the proposed method by training actual participants. We present our conclusions and future work in the final section.

II. Related work

There have been many research projects that support motion training [6][7] by analyzing user motions [8] using image processing [9], motion capturing [10], and wearable sensors [11]. Chan et al. [12] proposed a virtual reality (VR)-based dance training application using motion capture technologies in which the user wears a motion capture suit and follows the movements of a virtual teacher. The system provides several types of feedback on how to improve the movements, including displaying whether the position of each body segment is correct in real-time, showing a report on the user's performance after training, and showing a slow motion replay.

Hachimura et al. [13] proposed a motion training system using a motion capture system and mixed reality (MR) techniques. Their system uses a head-mounted display (HMD) that shows a CG character to visualize the movements of the trainer.

These methods focus on how to correct a user who has made an incorrect motion. However, users cannot imagine the situation where they are expert dancer using these methods. Our system provides the sense as if the beginner become an expert since he/she see the his/her expert performance in

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-177-9

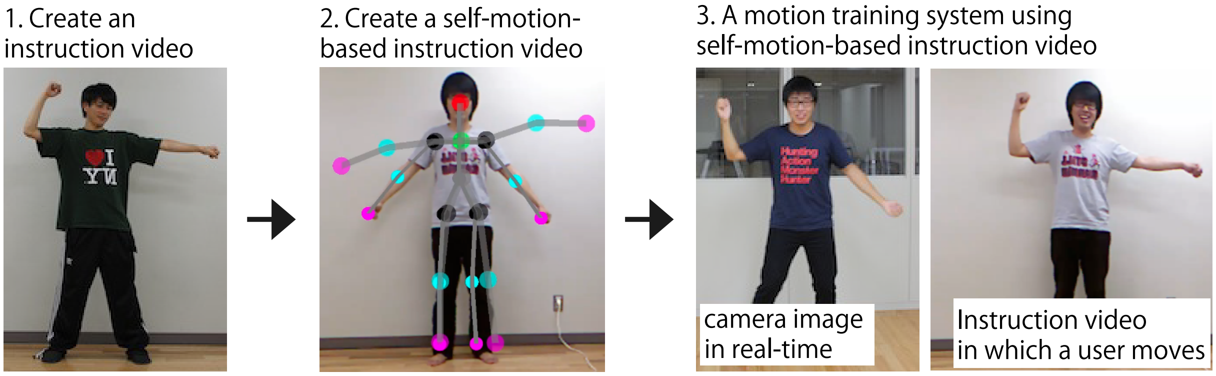
1. Create an instruction video
2. Create a self-motion-based instruction video
3. A motion training system using self-motion-based instruction video
camera image in real-time
Instruction video in which a user moves

Figure 1. Procedure of the proposed system

the movie. Our method can be combined with other methods to improve its overall effectiveness.

III. Proposed method

As a preliminary work, we developed a dance training system that increases user motivation by using a visual effect [14]. This system vibrates the user's image by using image processing to emphasize the intensity of their movement. As many individuals experimented with our system, we noticed that it was difficult for the average person to begin to enjoy dancing. We wanted to enable such people to enjoy dancing and to master basic dance techniques, and it occurred to us that it might be effective to show them video of themselves successfully dancing.

We performed a lot of trial-and-error experiments to achieve this purpose. However, it was difficult to create a video that realistically shows the user dancing like an expert dancer because the movements are so intense. Our main goal is to examine the meaning of and effect on people who watch a video of themselves dancing well, and we decided to use one simple motion to achieve this goal: *Stop*. This is a popping motion used in street dancing. *Stop* in which a dancer stops instantaneously - like a dancing robot - in accordance with the beat of music.

A. System Structure

Figure 1 shows the flow of our system, which consists of three subsystems. This flow is summarized below.

(1) The system records a video of an expert dancing to specified music while simultaneously recording the position of his/her body joints.

(2) It plays the recorded instruction video at a very slow speed, and the user moves in accordance with the video, including the position marks of the body joints, which are used as the materials for composing the instruction video.

(3) The user practices the dance motions by using the instruction video composed in Step (2).

We describe this process in the following sections in detail. In our prototype, we used a Kinect sensor [15] to acquire RGB and depth images. OpenFrameworks, OpenCV [16], OpenNI [17], and the NITE library were used to implement our prototype.

B. Recording the Instruction Video System

Our first subsystem records an instruction video that shows an expert dancing that includes RGB images, depth images, and BGM. It also records the trajectory of fifteen body joints, as shown in Figure 2(a). The frame rate for recording was 30 frames per second (fps).

C. Creating Self-Motion-Based Instruction Video

Our second subsystem creates a video that shows the user moving as if they are an expert dancer by using the instruction video recorded, as described in Section III.B. Figure 3 shows the flow of creating the self-motion-based instruction video. The procedure for creating the video is as follows.

(1) The user calibrates the position of his/her body joints to that of the dancer in the instruction video.

(2) The system plays the recorded instruction video at a very slow speed while simultaneously displaying the body joint markers.

(3) The user moves in accordance with the markers in the instruction video and the system records the user's images and the position of the body joints on each frame.

(4) The system calculates the distance between the position of the recorded user's body joints and that of the corresponding body joints on the instruction video on each frame and then determines the user's image for each frame of the instruction video.

This calibration process is described in detail in the following subsection.

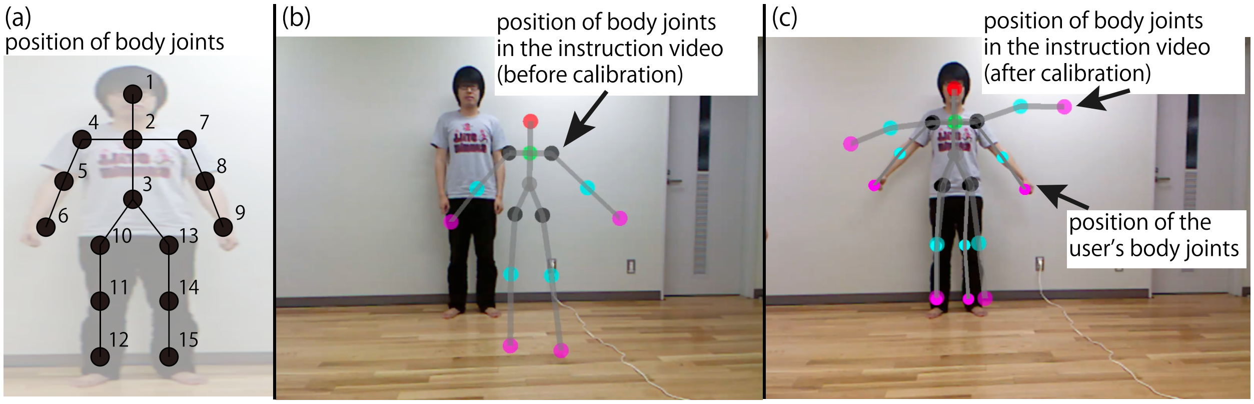

Figure 2. A calibration technique to position body joints

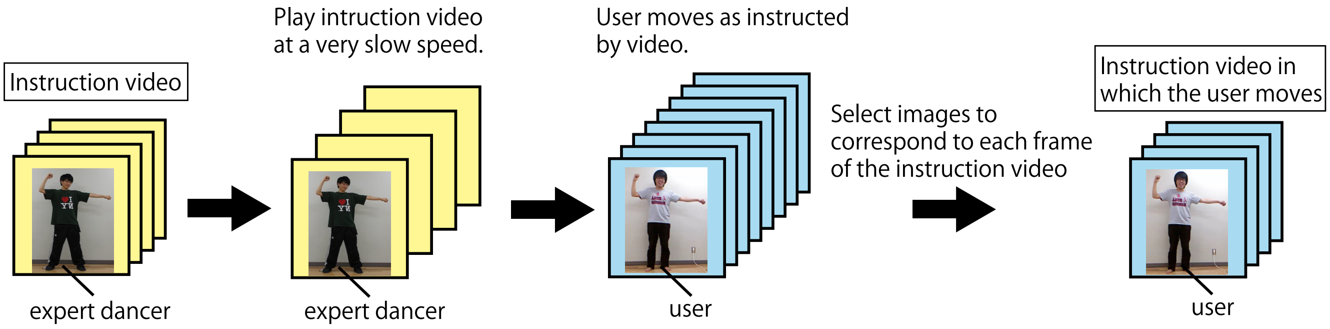

Figure 3. Procedure of creating a self-motion-based instruction video

1) Calibration of Body Joints: We use fifteen body joints, as shown in Figure 2(a). If the system displays the position of body joints from the instruction video without any modification, the user cannot move in accordance with it due to the difference between the size of the user's and dancer's bodies (Figure 2(b)). Therefore, our system calculates the distance between the user's body joints and the body joints on the instruction video and displays the body joints on the instruction video as adjusted to those of the user.

At a given number n, the system defines the body joints on the instruction video, $S_n(x,y)$, and the user's body joints, $U_n(x,y)$. The Euclidean distance $d(S_{n1}, S_{n2})$ and $d(U_{n1}, U_{n2})$ are calculated as:

$$d(S_{n1}, S_{n2}) = \sqrt{(S_{n1} - S_{n2})^2}$$
$$d(U_{n1}, U_{n2}) = \sqrt{(U_{n1} - U_{n2})^2}$$

The user can perform the same movements as the expert dancer by fitting the position of their joints to these on the instruction video, $S'_n(x,y)$, calculated as follows.

First, the system adjusts the center of the dancer's body joints, $S_3(x,y)$, to the center of the user's body joints, $U_3(x,y)$.

$$S'_3(x,y) = S_3(x,y) + \{U_3(x,y) - S_3(x,y)\} = U_3(x,y)$$

Next, the system adjusts $S_2(x,y)$ to $U_2(x,y)$ and calculates a ratio of length by using $d(S_2, S_3)$ and $d(U_2, U_3)$. In addition, because $S'_3(x,y)$ is adjusted to $U_3(x,y)$, $S'_2(x,y)$ needs to adapt to the difference $\{U_3(x,y) - S_3(x,y)\}$.

$$\begin{aligned} S'_2(x,y) &= S_2(x,y) + \{U_3(x,y) - S_3(x,y)\} \\ &+ (1 - \frac{d(U_2,U_3)}{d(S_2,S_3)})\{S_3(x,y) - S_2(x,y)\} \end{aligned}$$

$S'_4(x,y)$ is calculated by using the result of $S'_3(x,y)$ and $S'_2(x,y)$:

$$\begin{aligned} S'_4(x,y) &= S_4(x,y) + \{U_3(x,y) - S_3(x,y)\} \\ &+ (1 - \frac{d(U_2,U_3)}{d(S_2,S_3)})\{S_3(x,y) - S_2(x,y)\} \\ &+ (1 - \frac{d(U_2,U_4)}{d(S_2,S_4)})\{S_2(x,y) - S_4(x,y)\} \end{aligned}$$

In the same way, the system calculates all the positions of the body joints on the instruction video that have been adjusted to the user's body joints. As shown in Figure 2(c), the adjusted positions of the body joints on the instruction video overlap the user's body.

2) A Recording Technique Using Slow Instruction Video: The system creates a composite video showing the user dancing like an expert by recording the user's movements

ISBN: 978-1-61208-177-9

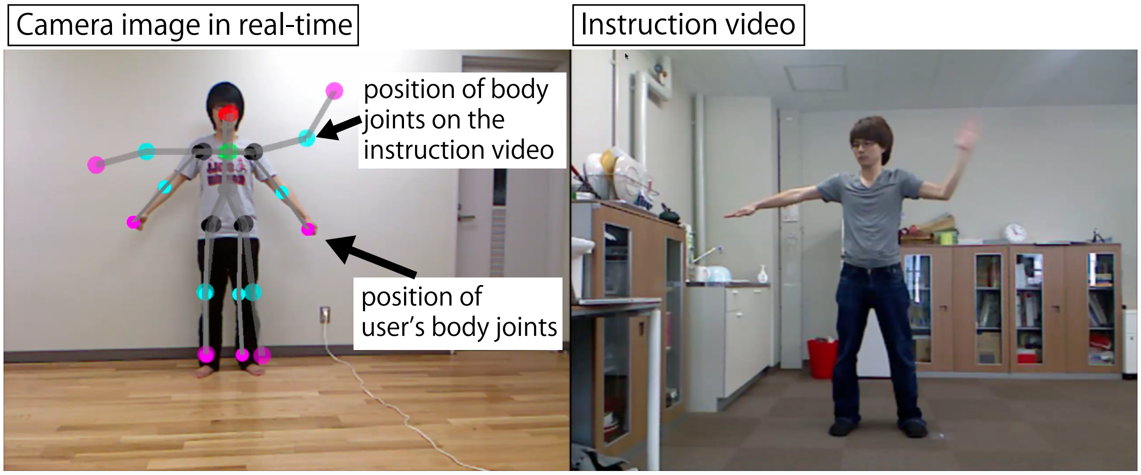
Camera image in real-time
Instruction video
position of body joints on the instruction video
position of user's body joints

Figure 4. Adjustment technique for body joints

in accordance with the instruction video. Obviously, it is difficult for beginners to dance like experts, so the system plays the instruction video at a slow speed, giving the user a longer time to get the movements right.

Specifically, the system plays the instruction video at a speed of one fifth and displays the adjusted position of body joints $S'_n(x, y)$. As shown in Figure 4, the system simultaneously displays the position of the user's body joints and the user follows the adjusted positions of the body joints on the instruction video. As the system plays the instruction video at the slower speed, it acquires the user's RGB image, depth image, and position of body joints at 20 fps. Therefore, when the system creates the composite video for 10 seconds, the user follows the instruction video for 50 seconds and the system records 1000 images at 20 fps. To summarize: the system creates a composite video showing the user dancing expertly from 1000 images.

3) Creating a Composite Self-Motion Image: As described in the previous section, the system records many images of the user, which it then uses to create the composite video. The sum of the Euclidean distance between the position of the body joints on the instruction video $S'_n(x, y)$ and the position of the user's body joints $U_n(x, y)$ is calculated as:

$$d(S', U) = \sqrt{\sum_{k=1}^{15}(S'_k - U_k)^2}$$

The system determines each frame of the composite video as the frame that has a minimum $d(S', U)$. By playing the images of the user that correspond to each frame of the instruction video at 30 fps, we can create a video that shows the user dancing like an expert.

D. Practicing with the Composite Self-Motion Image

The user can practice dance movements by viewing a real-time camera image on the left side and the instruction video on the right side, as shown in Figure 4. The user can switch between the instruction video and the composite video, which enables him/her to select the most effective video for whatever he/she is practicing. We also implemented a function that records a practice video and displays it on the left side, so the user can check his/her movements against the instruction video simultaneously.

IV. Evaluation

Our proposed system enables users to view an imitation of themselves dancing like an expert. In this section, we evaluate the effectiveness of the proposed system. The subjects used for each evaluation were 15 college students (including 1 expert dancer).

First, we created the dancer's instruction video. This video was 4 seconds long and featured a dancer performing *Stop* - in which the dancer pops like a robot in accordance with the beat - at 120 beats per minute. The movements included two types of motion: one for 250 ms at 750-ms intervals and the other for 250 ms at 250-ms intervals.

To create the composite video, subjects followed the instruction video at a speed of one fifth, as shown in Figure 4. Next, the subjects practiced the movements by following the instruction video for 3 minutes, as shown in Figure 5(a). Next, the subjects recorded their movements and checked them against the instruction video for 5 minutes, as shown in Figure 5(b). In each phase, the subjects were able to switch between their own video and that of the expert dancer. The subjects then answered the following three questions:

(1) If you practiced the movements by following the instruction video, which did you prefer, the instruc-

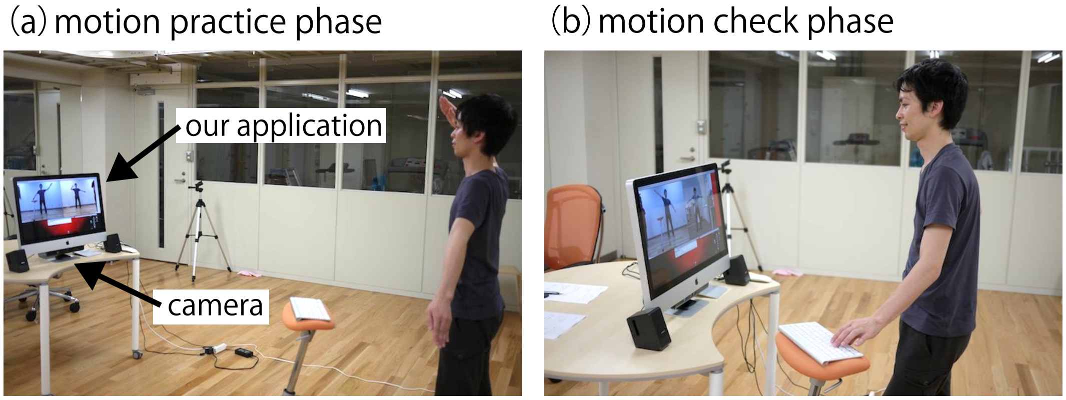

Figure 5. Snapshots of evaluation

Table I
EVALUATION RESULTS

	Number of subjects					The meaning of each evaluation
Evaluation score	1	2	3	4	5	
Evaluation (1)	3	3	3	5	1	1 (dancer's video) - 5 (user's composite video)
Evaluation (2)	4	1	3	3	4	1 (dancer's video) - 5 (user's composite video)
Evaluation (3)	0	0	0	5	10	1 (not interesting) - 5 (interesting)

tion video of the dancer or your own composite instruction video?

(2) If you recorded your movements and checked them against the instruction video, which did you prefer, the instruction video or the composite video?

(3) Do you think this system is interesting?

Table I lists the results of the evaluation. In questions (1) and (2), the opinions of the subjects are remarkably divided. We describe the evaluation and our conclusions in the following section.

A. Evaluation 1: Motion Practice Phase

The responses to question (1) are as follows.

- The composite video of myself was good. It reflected my movements more than the dancer's instruction video.
- I had a sense of rhythm by using the composite video. But I prefer the dancer's instruction video for practicing specific details of the movements.
- I was able to imagine myself successfully dancing by watching the composite video.
- I watched the dancer's instruction video more than the composite video of myself, because I felt that the instruction video was more accurate.
- First, I watched my composite video. Next, I watched the dancer's instruction video. This flow was effective for learning the movements.
- I disliked my composite video because it embarrassed me.
- I was skeptical of my composite video because I can't dance.

For subjects who were able to imagine themselves as dancers, it was effective to watch the composite video. Subjects who preferred the instruction video had the preconception that the instruction video was more correct than the composite video. Furthermore, they were not able to completely trust their own composite videos. Some subjects had different video preferences for each phase of learning the movements. We need to examine the role of timing in the use of the composite video.

B. Evaluation 2: Motion Check Phase

The responses to question (2) are as follows.

- It was helpful for me to study parts of my body by using the composite video.
- Using the composite video was helpful in terms of timing the movements.
- I was mortified when I was not able to dance well by using my composite video. This motivated me to try harder.
- By using the composite video, I could see my mistakes clearly.
- First, I watched my composite video. Next, I watched the dancer's instruction video. This flow was helpful for me to check the precision of my dancing.
- I preferred using the dancer's instruction video for checking my mistakes.
- Watching the dancer's instruction video made me feel that I have not yet mastered the dance.

In the phase of checking movements with the instruction video, subjects were able to compare the features of their bodies more than during the practicing phase. Subjects who preferred the dancer's instruction video used it to check the quality of their movements, while those who preferred the composite video used it to check the timing of their movements. In other words, preference for the dancer's instruction video versus the composite video differed depending on which movements were being learned. We intend to further evaluate our system to clarify the timing for using the composite instruction video in the future.

C. Evaluation 3: Impression of Subjects

The responses to question (3) are as follows.

- It was interesting to create a composite video by moving slowly.
- When I was able to dance well, my movements resembled my composite video. It was very interesting.
- By watching my composite video, I felt that I was able to dance well.
- My composite video was a little awkward.
- I felt a slight sense of incongruity in my composite video, but I still felt that it is effective for learning the movements.

By using our system, the subjects were able to enjoy learning dance movements. It seems that clear goals are effective when a beginner first learns movements. However, we need to improve our system to create a composite video of higher quality.

V. Conclusion

We proposed a system that enables beginners to practice dance movements by studying self-motion images that have been mapped onto an instruction video. Our system creates a composite video in which beginners appear to dance like experts. Experimental results showed that our system is effective for motion training in specific situations.

However, the motions that can be used are limited, and the composite video is not of high quality. We intend to improve the system in these respects in the future.

Acknowledgment

This research was supported in part by a Grant in aid for Precursory Research for Embryonic Science and Technology (PRESTO) from the Japan Science and Technology Agency and by a Grant-in-Aid for Scientific Research(A)(20240009) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

[1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake: Real-time human pose recognition in parts from single depth images, *Proceedings of Computer Vision and Pattern Recognition (CVPR'11)*, pp. 1297-1304, 2011.

[2] P. Haemaelaeinen, T. Ilmonen, J. HoeysniemI, M. Lindholm, and A. Nykaenen: Martial arts in artificial reality, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 781-790, 2005.

[3] U. Yang, and G. J. Kim: Just Follow Me: An Immersive VR-based motion tracking system, *Proceedings of the International Conference on Virtual Systems and Multimedia*, pp. 435-444, 1999.

[4] U. Yang, and G. J. Kim: Implementation and evaluation of "Just Follow Me": an Immersive, VR-based, motion tracking system, *Presence: Teleoperators and Virtual Environments, MIT Press*, Vol. 11, No. 3, pp. 304-323, 2002.

[5] D. Y. Kwon and M. Gross: Combining body sensors and visual sensors for motion training, *Proceedingsof the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pp. 94-101, 2005.

[6] J. N. Bailenson et al: The effect of interactivity on learning physical actions in virtual reality, *Media Psychology*, Vol. 11, pp. 354-376, 2008.

[7] S. Baek, S. Lee, and G. Kim: Motion retargeting and evaluation for VR-based training of free motions, *The Visual Computer*, Vol. 19, pp. 222-242, 2003.

[8] T. Shiratori, A. Nakazawa, and K. Ikeuchi: Rhythmic Motion Analysis using Motion Capture and Musical Information, *Proceedings of 2003 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems(MFI'03)*, pp. 89-94, 2011.

[9] L. Naveda and M. Leman: Representation of Samba dance gestures, using a multi-modal analysis approach, *Proceedings of Fifth International Conference on Enactive Interfaces*, pp. 68-74, 2008.

[10] C. Chua, N. H. Daly, V. Schaaf, and H. P. Camill: Training for physical tasks in virtual environments: Tai chi, *Proceedings of IEEE Virtual Reality 2003 Conference*, pp. 87-94,2003.

[11] K. Matsumura, T. Yamamoto, and T. Fujinami: Analysis of Samba Dance Using Wireless Accelerometer, *Nihon Kikai Gakkai Supotsu Kogaku Shinpojiumu, Shinpojiumu Hyuman, Dainamikusu Koen Ronbunshu*, Vol. 2006, pp. 216-221, 2006.

[12] J. Chan, H. Leung, J. Tang, and T. Komura: A virtual reality dance training system using motion capture technology, *IEEE Transactions on Learning Technologies*, Vol. 4, pp. 187-195, 2010.

[13] K. Hachimura, H. Kato, and H. Tamura: A Prototype Dance Training Support System with Motion Capture and Mixed Reality Technologies, *Proceedings IEEE Int'l Workshop Robot and Human Interactive Comm*, pp. 217-222, 2004.

[14] A/D Dance, http://ivrc.net/2010/en.html, 01.12.2011

[15] Microsoft XBOX Kinect, http://www.xbox.com/en-US/Kinect, 01.12.2011

[16] OpenCV http://opencv.jp/, 01.12.2011

[17] OpenNI, http://www.openni.org/, 01.12.2011