A Study on Taekwondo Training System using Hybrid Sensing Technique

Doo Young Kwon[+]

ABSTRACT

We present a Taekwondo training system using a hybrid sensing technique of a body sensor and a visual sensor. Using a body sensor (accelerometer), rotational and inertial motion data are captured which are important for Taekwondo motion detection and evaluation. A visual sensor (camera) captures and records the sequential images of the performance. Motion chunk is proposed to structuralize Taekwondo motions and design HMM (Hidden Markov Model) for motion recognition. Trainees can evaluates their trial motions numerically by computing the distance to the standard motion performed by a trainer. For motion training video, the real-time video images captured by a camera is overlayed with a visualized body sensor data so that users can see how the rotational and inertial motion data flow.

Key words: Taekwondo Training System, Motion Processing, Motion Training Video

1. INTRODUCTION

Watching and following motions performed by a trainer has been considered a fundamental principle for motion training. Beyond the conventional books and video, there have been plenty of interactive CD-ROMs with multimedia contents. Moreover, to provide a bidirectional motion based interaction, Virtual Reality (VR) systems have also been employed to check how a trainee follows an avatar [1,2]. Along this line, motion training environments have been investigated focusing on the interactivity based on human motion and feedback type. However, most of the current systems use only visual sensors to reconstruct the user's postures and check how the trainee imitates the trainer's motion. These approaches are limited in sensing and evaluating

detailed movements which can be critical in the practical motion training. Particularly, it is difficult for visual sensors to capture the rotary motions which are particularly important for Taekwondo movements. Furthermore, the previous systems have been developed only for trainees and not for trainers. Generally, traditional motion training is performed between trainers and trainees. A motion training system should be a medium where trainees practice following expert motions and trainers also perform motions to create instructive material which the trainees refer to. Thus providing functionality for both trainees and trainers is required.

In this paper, we describe our approach to build a Taekwondo training system for both trainers and trainees. As illustrated in Figure 1, Our goal is twofold. Firstly, we aim to combine body sensor and visual sensor data to provide an unique Taekwondo training method which improves the traditional motion training. Secondly, we provide intelligent functionalities of a Taekwondo training system that are important for trainers and trainees, including motion evaluation and motion

※ Corresponding Author : Doo Young Kwon , Address
: (157-030) 661 Deungchon, Gangsuh, Seoul, TEL :
+82-10-5003-514, FAX : +82-2-6393-3280, E-mail :
dykwon@kgit.ac.kr
Receipt date : Sep. 17, 2013, Approval date : Oct. 19, 2013
[+] Korean German Institute of Technology, Dept. of New
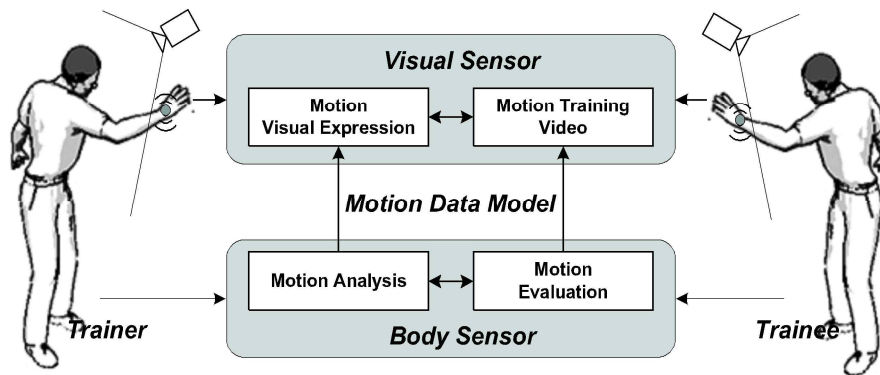Media, Associate Professor

Fig. 1. Concept diagram. Body sensors and visual sensors are combined to develop a motion training system for both a trainer and a trainee.

training videos.

The body sensor precisely measures the tilt detection, movement, and vibration of the body parts. On the other hand, like a mirror in conventional training places, the visual sensor provides the images of the users in real-time. Combining these heterogeneous sensor types, we improve the required tasks for motion training. For example, the accelerometer on a trainer's wrist provides precise tilt angles of the hand and amount of speed changes which are not visible to the naked eyes. Users can observe the sensor data of the performed motions and correct their motions by comparing them with another user's data such as a trainer.

We developed the motion chunk as a flexible segment unit to store a piece of Taekwondo motion information. Motion chunk allows us to make a computational motion model out of unstructured Taekwondo motions. With this model, we can apply motion detection and evaluation for motion training. We developed various functionalities for both trainers and trainees. Trainers and trainees can analyze their motion performances by watching a hybrid representation of visual and body sensor data and by generating a motion training video automatically. Thus, they do not have to manually record and edit the video for editing image frames.

2. RELATED WORK

A number of applications have been proposed for motion training systems. Davis developed a vision-based motion training system, called Virtual PAT (Personal Aerobics - Trainer) using IR light sources providing manually pre- recorded instructive videos and audio feedback [2]. Becker described a system for teaching Tai Chi gestures based on head and hand tracking by using a pair of stereo cameras [3]. Yang developed the "Just Follow Me" system using an optical motion capture system [2]. From this, Baek proposed evaluation methods by retargeting trainees' motion to the pre-generated avatar [4]. Chua developed a wireless VR system for Tai Chi training using a light-weight HMD display and optical motion capture device [1]. The trainees' motions are evaluated based on skeleton matching to measure how they mimic avatar motions. Takahata presented a martial art training method using sound generators and accelerometers without providing motion recognition and visual feedback [5].

Our system combines visual and body sensor data to develop a motion training system. While most of the previous training systems have been developed for trainees, ours supports both trainers and trainees. Using the machine learning techniques, we structuralize and label human motions in real time and automatically generate an

instructive motion training video. Thus, we achieved full automation while supporting motion training functionalities.

3. SYSTEM ARCHITECTURE

The system consists of a visual sensor (camera), a body sensor (wireless sensor network) and a display device (a projector or a monitor). The system is operated with a series of software components that constructs a motion data model by combining body and visual sensor data. Figure 2 describes the sequential data-flow between the components. During data acquisition, we collect signals from both the body sensor and the visual sensor. Wireless accelerometers transfer signals to the sensor base station which is connected to the main PC. We read the data with a sample rate of 10 Hz and transmit each packet 10 readings in size, so that the update frequency is overall 100 Hz. Synchronously, a webcam captures images and transmits the data to the host

PC.

The host PC accomplishes several steps to analyze the Taekwondo motion in real time. First, signal segmentation is performed to divide signals based on the structure of a motion chunk. To recognize reference motions from the segmented motion chunks, motion detection is performed based on HMMs [6]. Afterwards, the input motion is evaluated and assigned a score by comparing with reference data. For processing visual sensor data, first the acquired image data is processed in real-time to track the body position. Visual and body sensor data are synchronized by time-stamp. We generate visual feedback in the images incorporating the body sensor data.

4. MOTION PROCESSING

We developed various methods for motion processing which supports Taekwondo training functionalities in real-time. We explain how to decompose and analyze human motion. Then, we process the sensor data following the techniques outlined.
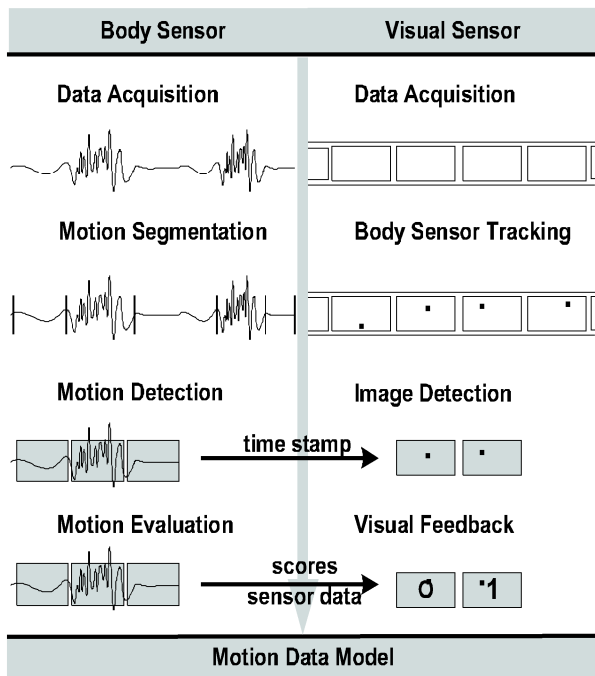
Fig. 2. The procedure to evaluate human motions and to generate visual feedback combining visual and body sensor data.
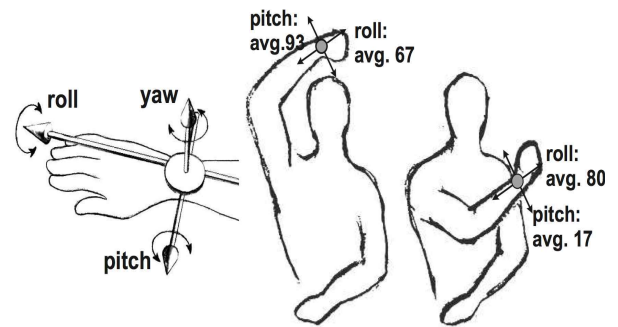
Fig. 3. Euler coordinate system of our body sensor. a posture with roll average 67 and pitch average 93, and a posture with roll average 80 and pitch average 17.

4.1 Motion Sensing

The motion is sensed through accelerometer as a body sensor. For the accelerometer we use an Euler coordinate system. The orientation is represented by three different angular values: yaw,

pitch and roll. These values are commonly used to describe the movement of a ship or a plane. To measure these values, the accelerometer is the most suitable because it enables the detection of tilt, movement, and vibration. The small size of such sensors is also appropriate for the human body. For example, we can attach the sensor to the wrist like a watch Figure 3. Then, the roll axis of the sensor is parallel to the forearm, and the pitch axis is horizontal and perpendicular to the roll axis. Yaw values point out the up- right position of the hand. Unlike the other two angles, yaw values are changed depending on the absolute orientation of the attached body part. In our setting, we utilize only 2-axes of the accelerometer providing pitch and roll. Using pitch and roll, we can measure the posture of the body part it is attached to, i.e. the forearm. We can infer other, adjacent body parts as well. For example, since the sensor is located on the forearm near the hand, it is also indicative for the orientation of the hand. That is, we can estimate whether the palm is facing back or front, or facing up and down.

4.2 Motion Chunk

We divide the body sensor information into two categories: *postures* and *gestures*. Postures are static expressions described orientation, whereas gestures are dynamic movements essentially defined by velocity and by the changes thereof. For instance, if the forearm rests in a certain position, it provides constant values for roll and pitch, i.e. a posture. The basic idea at the motion chunk is to decompose complex, sequential human motions into atomic units to simplify analysis. These units, called motion chunks, are similar in spirit to phonemes in speech recognition.

Following the definitions of postures and gestures we create two types of motion chunks: static chunks and dynamic chunks. We now define the motion chunk of one single motion as a combination of three chunks: start-static chunk+ dynamic chunk+end-static chunk. This is intuitively clear as it combines the start posture, the gesture, and the end posture. Figure 4 illustrates that a single step motion consists of two static chunks denoted by C and one dynamic chunk denoted by D. Likewise, two step motions are combined sharing one in-between static chunk and so forth.
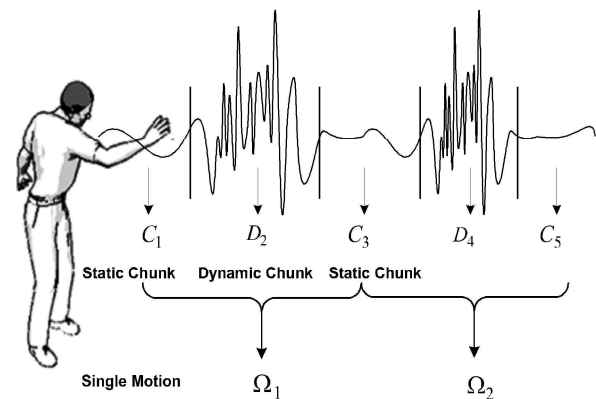

Fig. 4. Structure of a motion chunk. A recognized single motion consists of two static chunks and one dynamic chunk.

4.3 Motion Detection

We apply the concept of a motion chunk to represent the hidden states. The topology of each motion is represented as two distinct states which can be regarded as a two-state machine. We use the start-static chunk and the end-static chunk only, because the in-between dynamic chunk features a highly variant signal heavily depending on speed and power. Each HMM is created with the performance of two static postures (start and end). Using the signals as the observation sequences, we train the HMM parameters of each motion. We employ an iterative procedure called Baum- Welch method widely used to find a local maximum of the probability. Once the HMM model is trained, the system is able to detect the newly performed motion. For this, we employ the probability of the observations using a Viterbi algorithm. If the probability is high enough given

a HMM model of a motion type, we detect the input motion and generate a motion chunk with re-sampled data. The re-sampling is necessary to evaluate the quality of the motion. Then, the further processes, such as motion evaluation and motion training video generation, are executed.

4.4 Motion Evaluation

The gesture evaluation measures the similarity between two gestures using the similar techniques used in gesture recognition. The main task is to compare an input gesture to a template gesture that is performed by the same or different user, and determine how both performances are different. The result is used to improve user performance or to correct wrong gestures.

The evaluation process in our framework consists of both posture evaluation and gesture evaluation. We make the gesture instruction process similar to the practical motion training where users learn postures then perform gestures by connecting individual postures. Since posture and gesture co-occur, the coordination between postures is important for improving the efficiency of the gesture. The separation of postures and gestures helps users learn a complicated motion in a systematic way.

We compute three distinct scores for the start static chunk, the dynamic chunk, and the end static chunk each. The evaluation of two static chunks measures how the start and end postures have been performed respectively. The evaluation of the dynamic chunk expresses how the gesture is performed, with respect to power and speed. Currently, the result of evaluation is a simple numerical score. We compare the input gesture with multiple reference templates stored in our database and take the minimum distance as the final score. In practice, the minimum distance is better suited to measure quality than mean or median. The scores are normalized to a maximum value of 100 and displayed on the screen in real-time.

5. MOTION TRAINING VIDEO

A motion training video is necessary for trainees and trainers as a reference to follow and analyze preferred motions. However, producing such a video usually takes a lot of time. First, it requires simultaneous video recording during the trainer's performance. Also, the captured videos should be edited for the purpose of motion training such as selecting video frames and adding explanatory information. We provide a method for automatic generation of visual feedback. As soon as the input motion is detected, we save both the relevant video frames and the body sensor data. Then we generate a video displaying body sensor data along the tracked sensor positions.

5.1 Body Sensor Tracking

We extract sensor positions from the captured images and use the positions to generate visual feedback. We made various experiments to find suitable tracking solution for our purpose. First, the color band tracking highly depends on the training environment condition such as lighting and color. We also tested IR light sources, but they omit color information which is required. We found that color LED markers are most suitable for our purpose. Their brightness provides relatively robust tracking results in indoor training environments. We developed a simple vision tracking algorithm to find the pixel positions within a certain color and brightness range. The number and position of LEDs are designed depending on the sensor position. In our tests, we attached four LEDs at the four sides of the wrist bend Figure 5. This installation allows us to detect at least one point reliably even when the hand is rotated in different directions.

5.2 Visual Feedback

Visual feedback helps trainers and trainees to explain and improve their motion practice. Visual
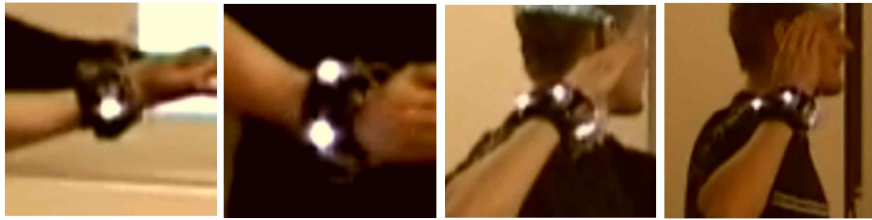
Fig. 5. Observations for body sensor tracking with LED markers on the wrist body sensor.
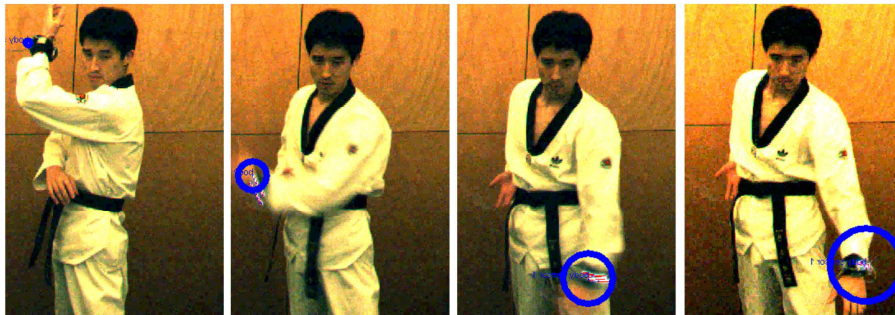

Fig. 6. A visual feedback of accelerometer sensor data in video images.

feedback for body sensor data is useful to show how the body sensor data is changing with the appearance of the posture. Especially for the trainees, visualizing motion path helps significantly to understand a dynamic gesture between two static postures. Thus we focus on visualizing body sensor data on the images along the motion path. We use the tracked sensor positions and design a simple template to display a moving circle along the path changing its size as a function of the magnitude of the acceleration Figure 6. There are various design alternatives, of course, varying the shape and its transformation rules.

6. CONCLUSION AND FUTURE WORK

We presented an approach to build a Taekwondo training system combining body and visual sensor data. We described a motion decomposition procedure called motion chunk for real-time motion analysis. Based on the motion chunk, we detect and evaluate a specific motion using Hidden Markov Models. We also presented an automatic video editing method to generate a motion training video including visualization of body sensor data. The system helps both trainers and trainees to improve fine static postures and dynamic gestures. So far, our research has mainly focused on analyzing single motions. Future work will be devoted to the analysis of longer sequences of motions. From this, users can eventually practice combinations of multiple motions.

REFERENCES

[1] C. Chua, N.H. Daly, V. Schaaf, and H.P. Camill, "Training for Physical Tasks in Virtual Environments: Tai chi," *Proc. IEEE Virtual Reality 2003 Conference*, pp. 87, 2003.

[2] U. Yang. G. Kim, "Implementation and evaluation of "just follow me": an immersive, VR-based, motion-training system," *Journal Presence: Teleoperators and Virtual Environments*, Vol. 11 Issue 3, pp. 304-323, 2002.

[3] D. Becker. Sensei, *A Real-time Recognition, Feedback, and Training System for T'ai Chi Gestures*, Master's thesis of Massachusetts Institute of Technology, 1997.

[4] S. Baek, S. Lee, and G. Kim, "Motion Retargeting and Evaluation for VR-based Training of Free Motions," *The Visual Computer*, Vol. 19, No. 4, pp. 222-242, 2003.

[5] M. Takahata, K. Shiraki, Y. Sakane, and Y. Takebayashi, "Sound Feedback for Powerful Karate Training," *Proc. International Conference on New Interfaces for Musical Expression*, pp. 13-18, 2004.

[6] D.H. Song, D.M. Kim, and C.W. Lee, "The Chinese Characters Learning Contents Based on Gesture Recognition using HMM Algorithm," *Journal of Korea Multimedia Society* Vol. 15, No. 8, pp. 1067-1074, 2012

Doo-Young Kwon

Mar. 2002~Feb. 2004: University of Washington, Dept. of Architecture, MS
Mar. 2004~July. 2007: ETH Zurich, Dept. of Computer Science, PhD

Feb. 2009~Current: KGIT. Dept. of New Media, Associate Professor
Research Interests: Media Art, Media Design, HCI, Virtual/Mixed Reality