JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 18, No. 8, pp. 1627-1633, Dec. 2017

인체 깊이 정보를 이용한 댄스 학습 시스템 개발

김 예 진
홍익대학교 게임학부

Development of Dance Learning System Using Human Depth Information

Yejin Kim
School of Games, Hongik University, Sejong 30016, Korea

[요 약]

인체의 댄스 학습은 실제로 수업을 참여하지 않고는 복잡하고 연속적인 신체 움직임으로 이루어진 전문가의 동작을 효과적으로 따라 하기는 어렵다. 본 논문에서는 인체 깊이 정보를 이용한 댄스 학습 시스템을 제안하고 있다. 제안한 시스템에서는 마커 프리 동작 갭쳐를 활용하여 댄스 전문가들로부터 다양한 예제 동작들을 갭쳐하고, 온라인 댄스 레슨에 사용하기 위해 동작 데이터베이스에 저장한다. 학생은 타블렛이나 키오스크 PC와 같은 사용자 단말기를 통해 데이터베이스에서 원하는 동작을 선택하고, 학습한 후 자신의 동작을 온라인 피드백을 받기 위해 강사에게 전송할 수 있다. 이 학습 과정에서 본 시스템은 네트워크 환경에서 학생과 강사가 효율적으로 동작 데이터를 교환하기 위해 빠르게 동작을 검색할 수 있는 방법과 다중 모드 화면을 제공한다. 실험 결과에 따르면 본 시스템은 학생들의 댄스 스킬을 주어진 시간 안에 향상시킬 수 있다.

[Abstract]

Human dance is difficult to learn since there is no effective way to imitate an expert’s motion, a sequence of complicated body movements, without taking an actual class. In this paper, we propose a dance learning system using human depth information. In the proposed system, a set of example motions are captured from various expert dancers through a marker-free motion capture and archived into a motion database server for online dance lessons. Given the end-user devices such as tablet and kiosk PCs, a student can learn a desired motion selected from the database and send one’s own motion to an instructor for online feedback. During this learning process, our system provides a posture-based motion search and multi-mode views to support the efficient exchange of motion data between the student and instructor under a networked environment. The experimental results demonstrate that our system is capable to improve the student’s dance skills over a given period of time.

색인어 : 댄스 학습, 예제 동작, 동작 캡쳐, 깊이 정보, 인체 동작

Key word : Dance Learning, Example Motion, Motion Capture, Depth Information, Human Motion

http://dx.doi.org/10.9728/dcs.2017.18.8.1627

Received 10 December 2017; **Revised** 26 December 2017
Accepted 28 December 2017

***Corresponding Author; Yejin Kim**

Tel: +82-44-860-2122
E-mail: yejkim@hongik.ac.kr

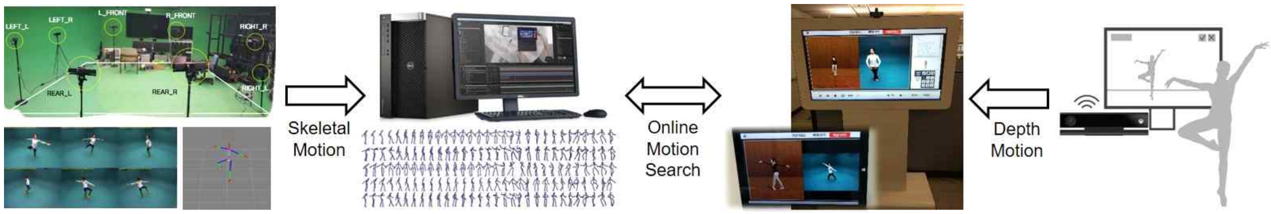


Fig. 1. Overall system for dance learning

Ⅰ. Introduction

Culture contents such as dance and song are attracting much attention worldwide as seen from the recent increase in popularity of Korean culture (i.e. Hallyu). According to UNESCO, culture contents belong to the intangible cultural heritage that become oblivious if not properly recorded and passed down to descendants [1, 2]. Over years, learning such culture contents has been mainly relied on a simple view on images and videos streamed from online web sites (i.e. Youtube and Google Video). In a such one-way learning process, a student lacks of the expert's feedback and has much difficulty in understanding key factors in the contents. This is especially true for the active contents such as dance, where a student should imitate complicated movements demonstrated from an instructor. For this reason, taking an actual class has been the most effective way to improve the dance skills based on the physical guidance of an instructor. However, some students simply do not have enough time to attend an actual class and to receive a proper feedback for dance guidance.

In this paper, we propose a dance learning system using human depth information. As seen in Figure 1, a set of example motions are captured from various expert dancers by using a marker-free motion capture, which consists of multiple RGB-D camera sensors. During the example motion capture, a set of RGB images at different viewpoints are reconstructed into 3D skeletal data and then archived into a motion database server for online lessons. Given the end-user devices such as tablet and kiosk PCs, the student can view and imitate a desired motion selected from the database. Using a RGB-D camera sensor attached to the device, the student can capture one's own motion and send it to an instructor through the online database. During this learning process, our system provides a posture-based motion search and multi-mode views to support an efficient exchange of motion data between the student and instructor under a networked environment. The experiment results demonstrate that our system is capable to improve the student's dance skills over a given period of time.

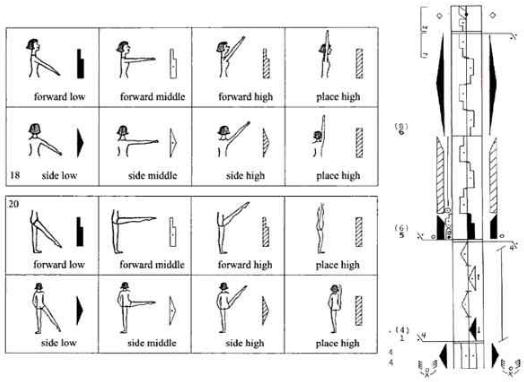
Ⅱ. Related Works

Learning dance motions without an instructor has long been suggested in different fields. As shown in Figure 2, different movement notes (i.e. Labanotation and EWMN) are used to express a sequence of body movements over time in the classical dance field [3, 4]. Later, these notations are digitalized into motion editing applications such as Isadora and DanceForms [5, 6]. However, these notes and applications require an extensive knowledge in human motion properties to edit human movements, which is not applicable to general users.

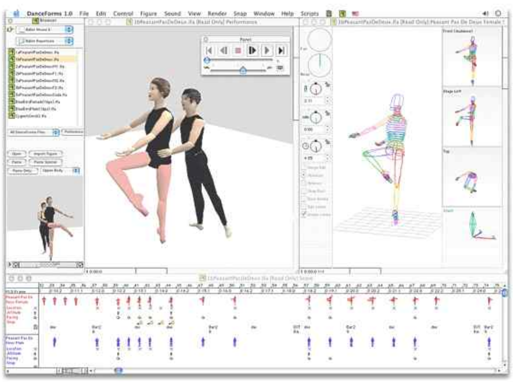
Based on the real-time motion capture technique, several video games such as Dance Central and Just Dance series, are introduced to imitate popular dances [7, 8]. In these systems, a single RGB-D camera sensor such as MS Kinect, is used to capture a player's motion [9]. During the game play, the captured motion is compared against the expert's motion stored in the game content for scores. However, these systems are mainly designed for an entertainment purpose and do not provide precise corrections on the user's motions. Recently, a virtual reality is added to the dance training system with the motion capture technique [10]. In this system, a student imitates an instructor's motion projected on the wall. However, the student should wear a special suit with a number of markers attached in order to imitate the expert motion. In addition, the overall system cost is expensive and requires a large space due to the optical motion capture system used; thus, it is not suitable for general students.

Ⅲ. Dance Learning System

3-1 Example Motion Acquisition



(a) Laban Movement Notations [2]



(b) DanceForms 2 by Credo Interactive [5]

(c) Dance Central Spotlight by Harmonix [6]

Fig. 2. Different systems used for dance learning

In our system, a set of example motions are captured from various expert dancers by using a marker-free motion capture, which allows a maximum freedom of movements during a dance performance. As shown in Figure 1, multiple RGB-D camera sensors (i.e. MS Kinect v2) are adopted to capture different dance types. For each camera, a sequence of color images along with corresponding depth data are retrieved at a different viewpoint. Based on the iterative closet point (ICP) method, a set of depth data are reconstructed into a single 3D skeletal motion by fusing a set of point clouds into one

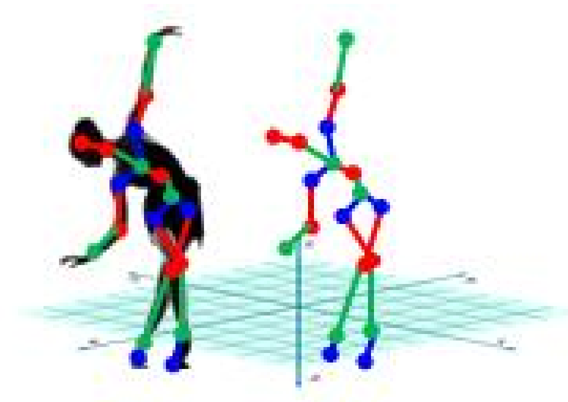


Fig. 3. 3D skeletal motion reconstructed for example motion

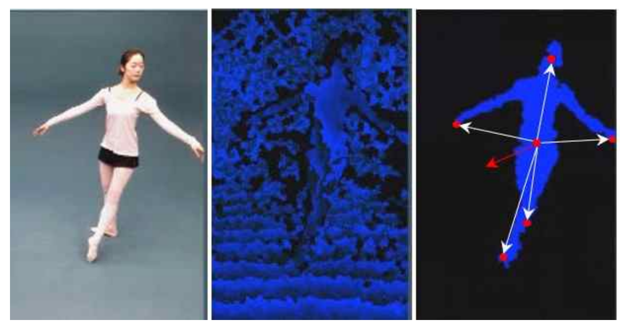


Fig. 4. User motion captured from a depth sensor: RGB image(left), depth image (middle), and filtered depth image with extreme points (red dots) and feature vectors (white and red arrows)

reference coordinate [11]. As shown in Figure 3, the reconstructed skeletal data are archived into the motion database with corresponding image sequences. These motion data are used as examples for online dance lessons, which can be accessed through the end-user devices. All the data in the database are segmented and labeled carefully by dance experts as a preprocessing step.

3-2 User Motion Abstraction and Comparison

Given the end-user devices, a student can browse and retrieve a desired motion from the database. Using a attached RGB-D camera (i.e. Kinect v2) on the device, the student can capture one's own motion and send it to an instructor for online feedback. While both example and user motions are archived into the database, its size grows fast, making the motion search a time-consuming job. Furthermore, the 3D skeleton reconstructed from the Kinect SDK is sensitive to the occlusion problem if not used with multiple cameras [12].

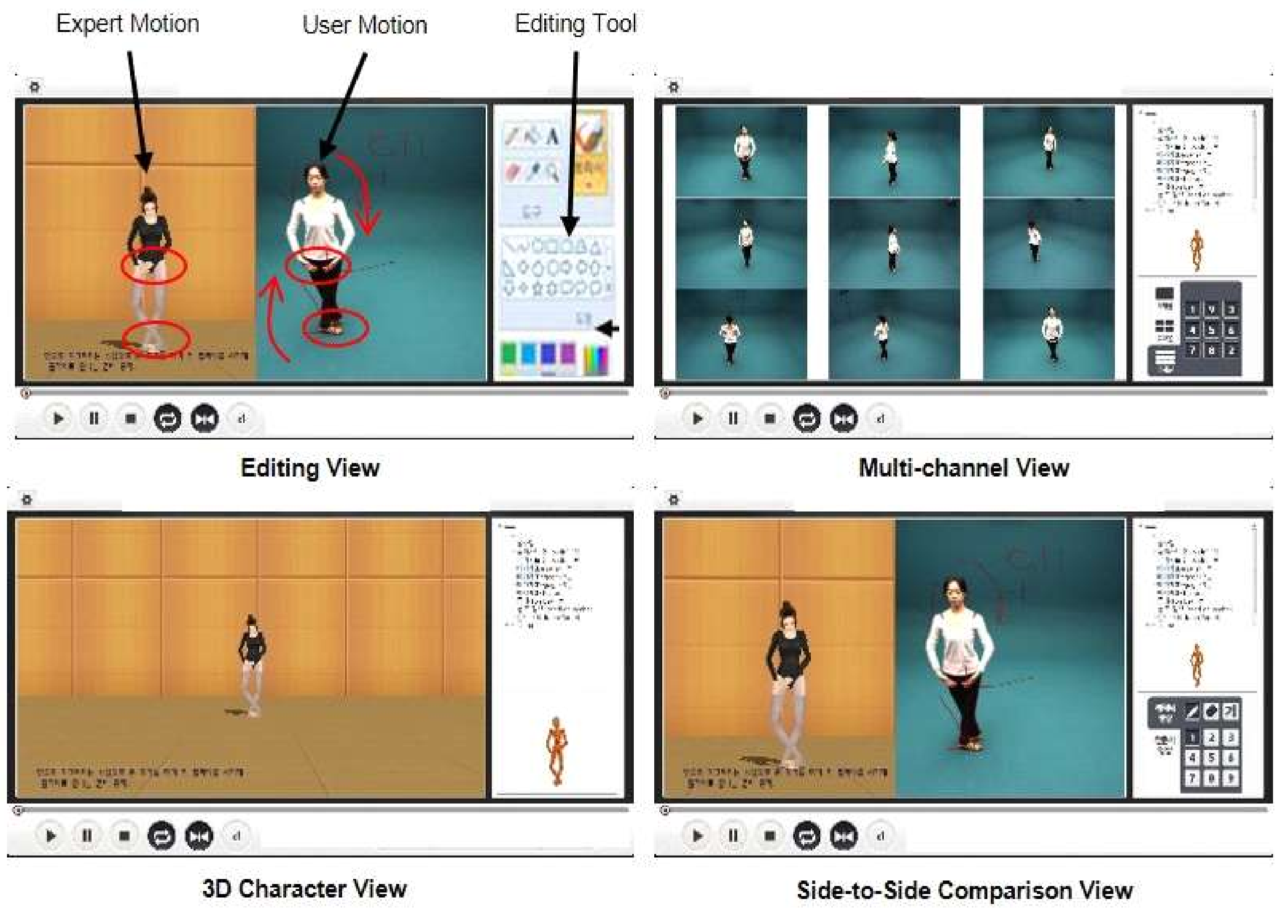


Fig. 5. Multi-view modes for online dance lessons

In our system, a user motion is abstracted by estimating a set of feature vectors. As seen in Figure 4, a depth image, which consists of a set of point clouds, is retrieved from a single depth sensor and filtered by subtracting a background noise. From the filtered depth image, a set of extreme points, p_i, where $i \in$[*Center, Head, HandRight, HandLeft, AnkleRight, AnkleLeft*], are detected from the human body parts. For an efficient search for p_i, a quadtree structure is applied to limit the search area, while histogram of oriented gradients (HOG) and support vector machine (SVM) are used to specify p_i from the segmented body parts [13].

Given p_i, a set of feature vectors, $\vec{v}_j$, where $j \in$[*Head, HandRight, HandLeft, AnkleRight, AnkleLeft*], can be estimated between the extreme points while a body orientation vector, $\vec{n}$, is estimated at p_{Center}. For a posture comparison between a frame in an example motion, P_A, and a frame in a user motion, P_B, their similarity is measured as a minium sum of weighted angular differences:

$$D(P_A, P_B) = \min(D(\vec{v}_A, \vec{v}_B) + D(\vec{n}_A, \vec{n}_B) \tag{1}$$

where

$$D(\vec{v}_A, \vec{v}_B) = \sum_{j=1}^{5} w_j \| \vec{v}_{A,j} - T_{\theta, \Delta x, \Delta z} \vec{v}_{B,j} \|^2 \tag{2}$$

and

$$D(\vec{n}_A, \vec{n}_B) = w_{Center} \| \vec{n}_{A,Center} - \vec{n}_{B,Center} \|^2 . \tag{3}$$

Here, w_j is a weight value that sets the importance of $\vec{v}_j$ during the search. Futhermore, $T_{\theta, \Delta x, \Delta z}$ is used to align $\vec{v}_{B,j}$ to $\vec{v}_{A,j}$, which rotates $\vec{v}_{B,j}$ about the vertical y-axis and translates $\Delta x = x_{A,j} - x_{B,j}$ and $\Delta z = z_{A,j} - z_{B,j}$ for a

Table 1. Dance motion database: All example motion are archived at 30 fps.

Category	Types	Total Frames (s)
Ballet	25	255,210 (8507)
K-pop	10	63,660 (2122)
Traditional Korean	10	63,330 (2111)

more precise comparison.

Given a single posture captured from the user, our system lists similar postures in the similarity order. To avoid abundant search results, a window size of N_f frames are used to skip the similar postures in neighbor frames.

3-3 Online Dance Lesson

In our system, a student and an instructor exchanges motion data (i.e. image sequences and 3D skeletal motion) through the motion database server using the end-user devices. As seen in Figure 5, the student can browse the database and selects a desired motion for an online lesson. For this learning process, our system provides different view modes: the multi-channel view for displaying an example motion at different angles, the 3D character view for displaying an example motion in an articulated polygonal model, and the side-to-side comparison view for displaying an expert and a user motion at the same time.

Using the attached camera sensor on the device, the student can capture and send one's own dance movements to an instructor for online feedback. Using another device, the instructor can review the captured motion and send it back to the user through the online database server. For the instructor, an editing view is provided to allow correction marks on the frames in the input motion sequence. It is noteworthy that this online interaction can be made in real time if the devices are connected in broadband connections between the student and instructor.

Ⅳ. Experimental Results

For the motion database, various example motions were captured from ballet, K-pop, and traditional Korean dancers. Table 1 shows different types and the total frames (seconds) captured for each of the dance category.

To demonstrate the effectiveness of our system, a total of 10

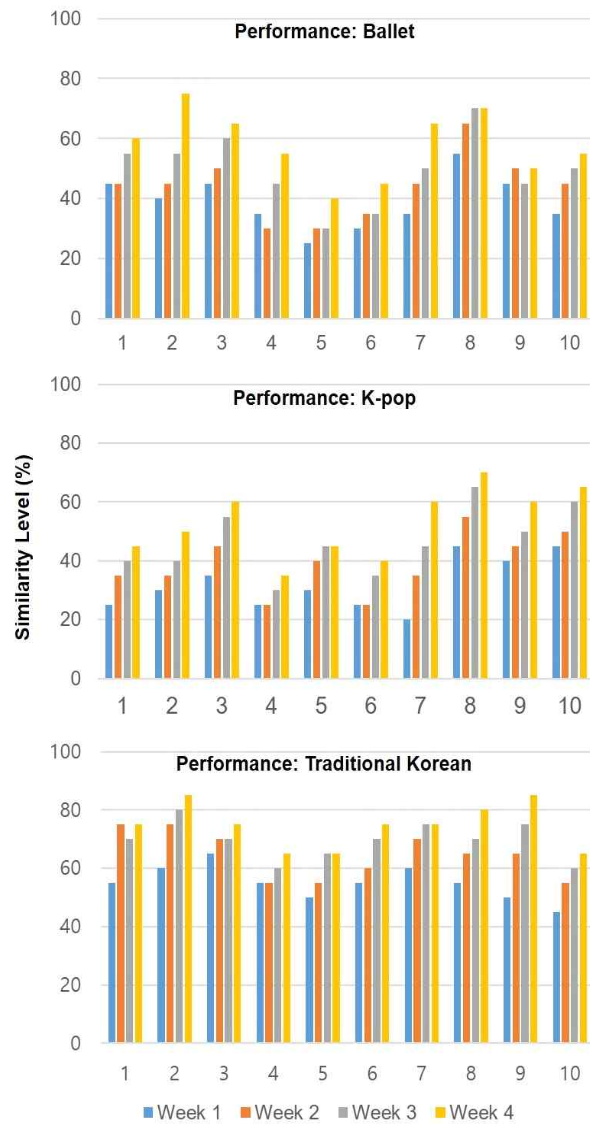


Fig. 6. Dance learning performance of participants over weeks

students, who had not formerly taken any dance lesson, participated in online dance lessons such as ballet, K-pop, and traditional Korean, using our system. Over four weeks, each student has taken the learning session 3 to 4 times a week for each dance category and received online feedbacks from an instructor. At the end of each week, the instructor scores the performance level for each student by counting the number of incorrect key postures performed by the students. Figure 6 shows that most of the participants have improved their performance levels over the given period of time. It is noteworthy that the increase in their performance levels differs from one dance category to another. For example, most participants have felt the

K-pop dance is the most difficult to learn while the traditional Korean dance is least difficult from the online lessons.

Ⅴ. Conclusion

The proposed system is designed to provide online dance lessons for a student who is unable to attend an actual class. For the motion database, a set of example motions are captured from various expert dancers via the marker-free motion capture. Using the end-user devices such as table and kiosk PCs, the student can select and practice a desired motion from the database while an instructor can review the motions sent from the student under a networked environment. During this learning process, our system provides the posture-based motion search and multi-view modes to support the efficient exchange of motion data between the student and instructor. The experimental results demonstrate that our system is capable to improve the student's dance skills over a given period of time.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2017R1C1B5017000)

References

[1] United Namtions Educational, Scientific and Cultural Organization (UNESCO). Intangible cultural heritage [Internet]. Available: https://ich.unesco.org/en/home/.

[2] S.-H. Lee and A.-R. Mo, "A Study on the Use of Sound Source in Music Education for Children," *Journal of Digital Contents Society*, Vol. 18, No. 3, pp. 509-516, June 2017.

[3] Dance Notation Bureau. Labanotation [Internet]. Available: http://www.dancenotation.org/.

[4] The Noa Eshkol Foundation for Movement Notation. Eshkol-Wachman Movement Notation (EWMN) [Internet]. http://noaeshkol.org/.

[5] TroikaTronicx. Isadora [Internet]. Available: https://troikatronix.com/.

[6] Credo Interactive. DanceForms 2 [Internet]. Available: http://charactermotion.com/products/danceforms/.

[7] Harmonix. Dance Central Spotlight [Internet]. Available: http://www.harmonixmusic.com/games/dance-central/.

[8] Ubisoft Entertainment. Just Dance Now [Internet]. Available: https://just-dance.ubisoft.com/en-us/home/.

[9] Microsoft. Kinect Camera Sensor [Internet]. Available: https://developer.microsoft.com/en-us/windows/kinect/.

[10] 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, No. 2, pp. 187-195, April-June 2011.

[11] Y. Kim, S. Baek, and B.-C. Bae, "Motion Capture of the Human Body," *ETRI Journal*, Vol. 39, No. 2, pp. 181-190, April 2017.

[12] A. Kitsikidis, K. Dimitropoulos, S. Douka, and N. Grammalidis, "Dance analysis using multiple kinect sensors," in *Proceedings of the international conference on computer vision theory and applications*, Lisbon, pp. 789-795.

[13] S. Hong and M. Kim, "A Framework for Human Body Parts Detection in RGB-D Image," *Journal of Korea Multimedia Society,* Vol. 19, No. 12, pp. 1927-1935, December 2016.

Yejin Kim

2000 : University of Michigan, Ann Arbor (Bachelor in Computer Engineering)
2003 : Korea Advanced Institute of Technology (M.S. in Computer Sciene)
2013 : University of California, Davis (Ph.D. in Computer Sciene)

2003~2007: Researcher, Electronics and Telecommunications Research Institute (ETRI)
2013~2016: Senior Researcher, Electronics and Telecommunications Research Institute (ETRI)
2016~Current: Assistant Professor, School of Games, Hongik University
※Fileds of Interests : Human Motion, 3D Character Animation, Computer Graphics, Serious Games