Multimed Tools Appl
https://doi.org/10.1007/s11042-018-5922-8

A virtual tutor movement learning system in eLearning

Hsin-Hung Chiang[1] **· Wei-Ming Chen**[2] **·**
Han-Chieh Chao[1] **· De-Li Tsai**[2]

Received: 9 November 2017 / Revised: 13 February 2018 / Accepted: 21 March 2018

Abstract This paper provides a training system with an augmented reality interactive body movement for movement learning, such as gymnastics, martial arts, sports or dance learners. The technology of depth image sensor is used to detect, track and measure the body's movements to collect the path of body's movement in 3D space, and all images has been further modified to reveal the function of feedback immediately. The learner follows up the pre-recorded tutor's movement to imitate tutor's movement step by step. The training system would judge whether the learner's movement correct or not, compares with tutor's, and offer an analysis result in-situ. The learner could get a training as well as real expert guides without any constraints of space and time and with low cost in this system.

Keywords eLearning · Motion capture · Depth sensing

1 Introduction

The use of technology in virtual tutor has been developed to train the learner of gymnastics, martial arts, sports and dance. The function of immediate feedback in virtual tutor system

✉ Hsin-Hung Chiang
d9823008@gms.ndhu.edu.tw

Wei-Ming Chen
wmchen88@gmail.com

Han-Chieh Chao
hcc@mail.ndhu.edu.tw

De-Li Tsai
a5973149@gmail.com

[1] Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, Republic of China

[2] Department of Computer Science and Information Engineering, National Ilan University, Yilan, Taiwan, Republic of China

would be similar with the traditional face-to face teaching, provide a training behind the constraint of space and time. It is difference from real expert guides, a traditional face-to-face teaching; and also difference from the video teaching and on-line teaching.

There are many researches have recognized that eLearning is one of important method for education [5, 8, 10]. The two models—synchronous and asynchronous, is sorted in eLearning. The synchronous eLearning, involves online studies, through chat and videoconferencing where teacher guides and discusses with learner immediately. Both of teacher and learner have to be online at the same time. In the other hand, the asynchronous eLearning always takes digital teaching or multimedia as tool for teaching, learner would begin ones learning when learner is available. However, the learner cannot correct one's movement from the immediately feedback.

Both of two eLearning models provide the feedback by pre-set questionnaire, based on the static knowledge, for learner to input text, images or video playback in some device. It would be not an in-situ feedback. In other word, the feedback would not be effectively assessed. The number of learners is also a limitation in synchronized on-line learning, that also limits the learners to get the best teaching qualities. The other choice for learner is to watching non-synchronized media repeatedly.

In past studies, there have been many studies using Kinect to score dance performance [1, 6]. It seems to be difficult for teacher to use the existing Dataset for virtual imagery to record video for eLearning applications [1]. Some of applications have to prepare the pre-generated sample poses which can only be applied to some special application [6].

This article extents efforts to enhance the efficiency for learner using eLearning. The novel set of learning model is, for learner to learn one's body movements, with immediately feedback, and without constraint of space and time. The virtual tutor images in system provides observations and captures the learner's body movement in three-dimensional space, and then analysis the accuracy of the learner's body movement in-situ. It would then correct the learner's movement immediately. This virtual tutor system, is helpful for a learner to improve the body movement, would be available to replace the current traditional face-to-face teaching and other eLearning system with low cost and high efficiently.

2 Related work

The basis of this study is outlined for the virtual tutor movement learning. It describes the image capturing and algorithm for gesture segmentation and recognition in this study.

2.1 Microsoft Kinect

The depth information of captured video is one of important target that researcher studies the human post recognition in recent years [9]. It will cost researcher much time to catch the depth information from captured video to calculate each frame data from video, to say nothing of mentioning to capture the real-time video. Fortunately, Microsoft Kinect provides the function of depth information with a simply process of image [1, 6, 7]. The SDK in Kinect further provides the various functions for processing a number images. The Microsoft Kinect is suitable tool for our study.

2.2 Douglas-Peucker algorithm

It is recognized that delivers the best perceptual representations of the original lines for the classical Douglas-Peucker line-simplification algorithm [3]. The algorithm is used extensively for both computer graphics and geographic information systems. It is suitable for us to simplify paths for tutor's and learner's.

The use of Douglas-Peucker algorithm is as following steps, as shown in Fig. 1:

(1) Set a threshold value {th}, and keep all points that distances from line are longer than {th}.
(2) Find two farthest points {A, B} and make a straight line $\{\overline{AB}\}$. Calculating the distance of each other points from $\{\overline{AB}\}$ to find the longest distance of them. If the distance of point {C} is the longest and longer than {th}, this point {C} will be kept.
(3) Taking the point {C} to produce the two lines $\{\overline{AC}, \overline{BC}\}$ and repeating the step 2 to find the longest distance of each other points based on two lines $\{\overline{AC}, \overline{BC}\}$, respectively. Finding the longest distance of point {D}, and make sure it is longer than {th}, this point {D} will be kept.
(4) Repeating step (3) until the distance of point is not longer than {th}.
(5) Finally, connecting all kept points to obtain a simplified line.

2.3 Integral image

The integral image [2] is a quick and effective method to calculate the sum of gray values. The average intensity of given image would be calculated throughout the Eq. 1. Figure 2 shows the process of simplified and grayscalized image by using the integral image. The process of integral image is used to accumulate of values from left to right and top to bottom. Figure 2a shows the distribution of original gray value. The value of area D in Fig. 2b is obtained via ABCD – (AC + AB) + A. It is easily and quickly to calculate every gray value from original image, i.e. part of ABCD, AC, AB or A, as shown in Fig. 2b. The gray value in red box would be accumulated from integral image, as shown in Fig. 2c.

$$Integral\left(i_x, i_y\right) = \sum_{x=0}^{x} \sum_{y=0}^{y} image(x, y) \tag{1}$$

Where (i_x, i_y) is gray value in coordinate (x, y)
Image (x, y) is the sum of gray scale of the area (x, y)

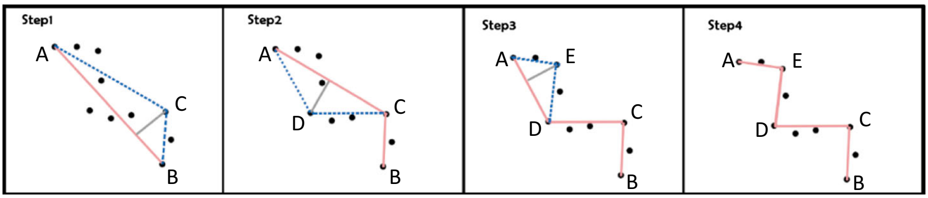


Fig. 1 Process of simplified line by Douglas - Puke algorithm

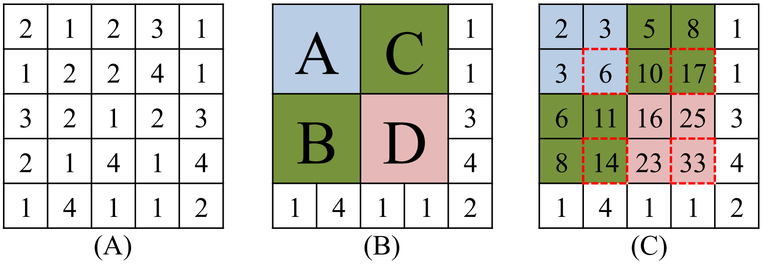


Fig. 2 The process of simplified and grayscalized image from integral image

2.4 Haar-like feature

The Haar-like feature is one of the most popular methods for face detection [4, 11]. It uses the integral image from gray value of original images to identify the human feature by comparing with special pattern, as shown in Fig. 3. The dark part in Fig. 3 is the feature of special pattern that used for detecting in this study. The gray value would overcome the effect of the different color skin. The method of Haar-like feature affords a fast and easy method to implement without any hesitate.

3 System architecture

We design a set of interactive learning system for teaching body movements in the three-dimensional space. The virtual tutor image exhibits a correct micro motion for learner to observe and imitate, the system would compare the learner's movement with virtual tutor simultaneously, as shown in Fig. 4. Each body movement will generate a path in the three-dimensional space, learners need try to figure out the body movement. When the learner moves his limbs, this system will compare with virtual tutor and feedback the comparison results in-situ. The system would show the comparing path and correct rate and re-practice and advise learner to improve immediately.

The architecture of our system includes three parts: Video Process (Kinect sensor), Comparison and Interactive display.

3.1 Part A. Video Process

Figure 5 shows the flowchart of video process. The tutor would create some the standard body movements with various revisions for different degree in database, including virtual tutor video and standard movement paths in three-dimensional space. Learner download the revision which

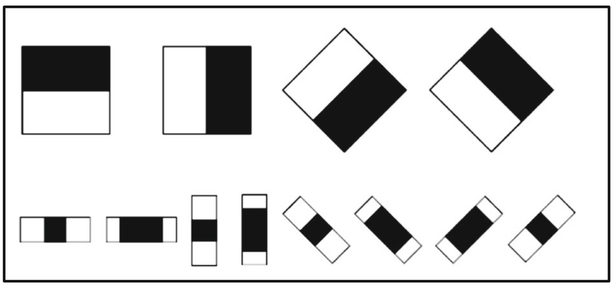
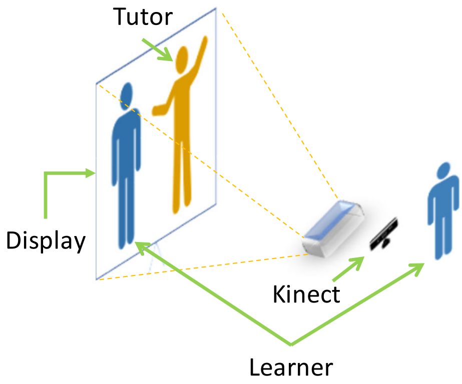


Fig. 3 Some common patterns used in Haar-like feature method

Fig. 4 Map of the scene layout



is suitable for learner's degree to learn the body movement. The subsequent comparison is generated when the learner began to use this system.

3.1.1 Teacher data

After recorded, the video of whole action is divided into multiple, continuous small pieces, and then editing those pieces in post-production.

3.1.2 Removing backgrounds

The depth image and color image are obtained from Microsoft's Kinect depth sensor and camera, as shown in Fig. 6. The Fig. 6a, b, c and d shows the depth image, color image, body area and color image of body, respectively. It has to avoid presenting any obstructions between learner and the Kinect. In other word, learner is the only event which is the closest to Kinect. It would be convenience for senor to distinguish the difference between human body and scene according to the minimum depth value. However, the minimum value would cannot avoid producing the interference of noise. There are two steps as shown in following would be used for further confirming the body shape.

a. Find body area

   The smallest depth of image is taken as standard, the part of image where the depth value is the closest to the standard value would be considered as human body. The tool of integral image equation is used to accumulate the values from left to right and top to bottom, to identify the body area.

   Figure 7 shows the depth map of depth conversion using Kinect directly. The color in

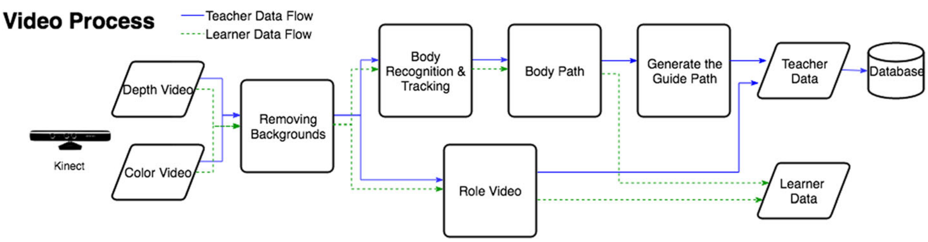


Fig. 5 Video process flowchart

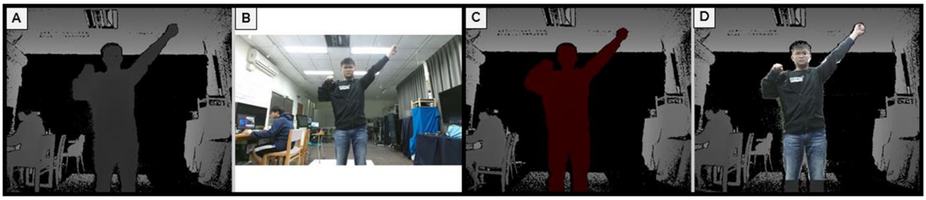


Fig. 6 **a** Depth image. **b** Color image. **c** Body area. **d** Color image of body

image depends on the distance — the darker color represents the closer distance. The red marked area, treated as invalid area, represents Kinect cannot determine the distance, would be ruled out in the subsequent calculations. The error of body area would be further made up by the following processing.

It is necessary to find the average depth value of effective point of original image. Figure 8a represents the original image, it consisted of valid area with different gray value and invalid area. The gray value of invalid area in original image is defaulted as "0". The total depth value of image would be obtained, as shown the example of depth map in Fig. 8 (B). On the other way, the gray of valid area in original image is set as "1", the invalid area is defaulted as "0". The total of effective point would be obtained, as shown the example of depth map in Fig. 8c. The thresholds, the average depth value divided by the no. of effective points, would reduce the error significantly caused by the invalid area. Furthermore, input the estimation of learner's height and movement range in Kinect. After the process that mentioned above would result in the emergence of complete body area.

b. Displaying the color image of body

The color image obtained from the previous process has to be further modified before be used. Because of the position of the camera and the sensor in Kinect is not on the same viewpoint. The conversion function of Kinect provides the corrected the different coordinate between the depth image and the color image. A normal color image is displayed in screen after Kinect corrected.

Fig. 7 The depth image offered by the Kinect directly

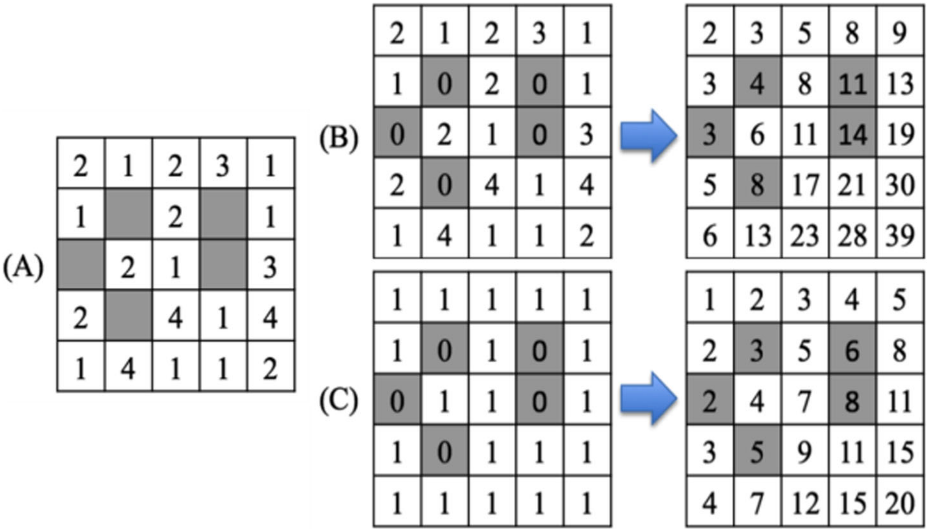


Fig. 8 **a** Original image. **b** Total depth value of image. **c** Average depth value of effective point

3.1.3 Body recognition & tracking

The Kinect provides the recognition of body skeleton, as shown in Fig. 9. There are some modifications based on Kinect have been designed in this study to ensure the skeleton is more stable in the system.

a. Identifying front side and back side

   The body's area obtained from the previous step. The Haar-like feature method is used to identify the front side by the color image face detection. It would compensate the Kinect that detects the depth to recognize the human body skeleton but not the front or back side.

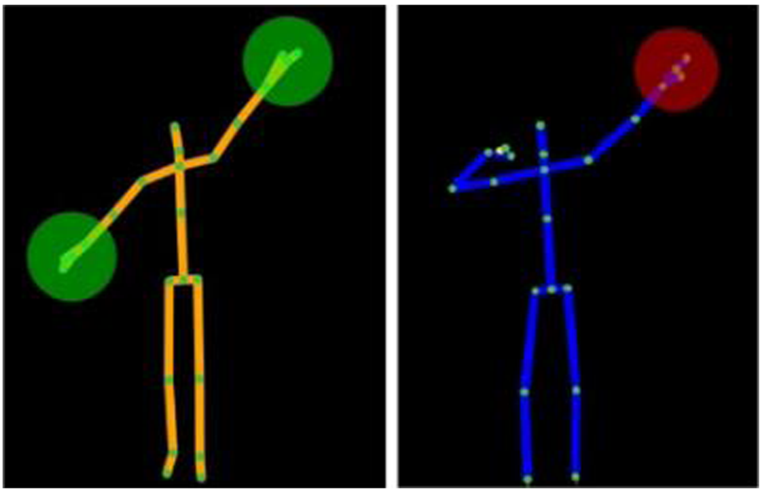


Fig. 9 Skeleton recognition by Microsoft Kinect's SDK

b. Record and adjust the size of the skeleton

Guide path of standard movements would display in screen. The size of learner body is always not as well as standard. Take the height for example, the different height would lead to the learner can't achieve the standard movement or be wrong movement from Kinect detects, as shown in Fig. 10a where uncorrected path detected by Kinect directly.. The guide path has to adjust based on the different body of learners, i.e. height, length of arm or foot, as shown in Fig. 11. The correcting process is described as followings:

(1) Assuming the position of the Kinect is the origin of the spatial coordinates, it would be calculated by using the projection method.
(2) Each height records (H_1 and H_2) and the distance records (D_1 and D_2) has the same proportion, that is used as adjustment ratio to modify the entire path record to scale the skeleton.
(3) After scaled, the entry path records would match the learner based on the position of the body center point.

The corrected path is obtained after adjusted by the proportion, as shown in Fig. 10b.

c. Body path

In the previous step, the body tracking record should be modified further because the distance from Kinect is always uncertainly for the learner, as shown in Fig. 12. The Kinect SDK provides the conversion function to correct the coordinate positions and depth value in different distance.

d. Simplifying the body path

It is not avoided that the jitter of path would be generated caused from limitation of hardware and/or conversion equations. That makes the display is difficult, and lead to some errors for learner to learn. The method of Douglas-Peucker algorithm is used to simplify the path by selecting the main point but keep all accepted points.

The threshold value is very important for the Douglas-Peucker algorithm, it is necessary to build a mechanism to ensure the best setting of it. The threshold value should be further test by Eq. 2. where calculates the difference between the original and simplified line. The threshold value has to be adjusted when the different value is out of expected range. After fully tests, an optimum threshold value would be used for each learner. The Eq. 2 is shown as followings:

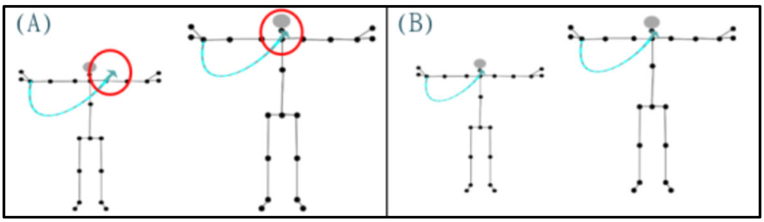


Fig. 10 **a** Uncorrected path detected by Kinect directly. **b** Corrected path after adjusted

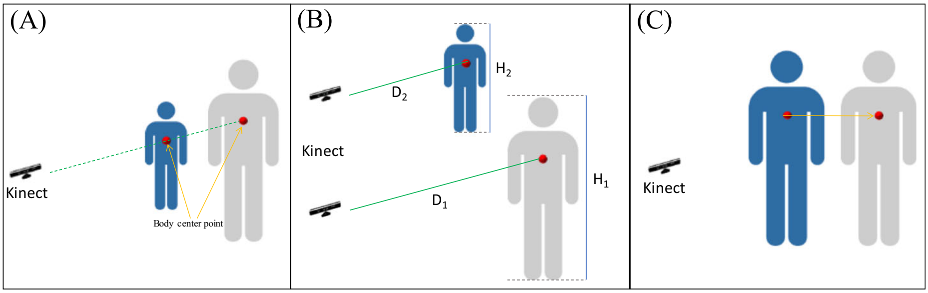


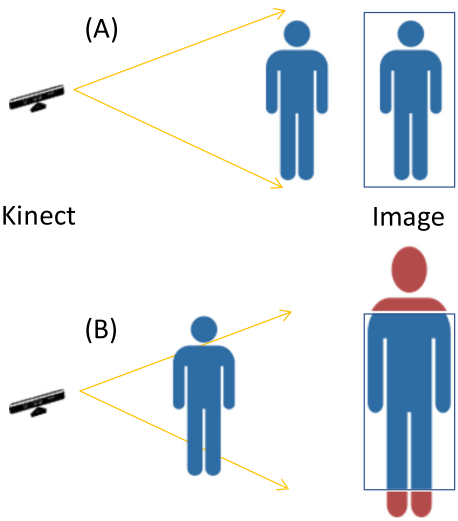
Fig. 11 The process of corrected path

$$V_{diff} = \frac{\sum_{i=1}^{n} \left| \frac{D_{min(S_i, L(S,th))}}{th} \right|}{n} \tag{2}$$

th	Threshold.
S	Original path data.
S_i	The ith point in the Original path data. ($i = 1$ *to* n)
$L(S, th)$	The simplified line via the threshold.
$D_{min}(S_i, L(S, th))^2$	The distance between ith point to $L(S, th)$. ($i = 1$ *to* n)
V_{diff}	The degree of difference after the path was simplified.

Figure 13a shows the original path obtained from Kinect. It is found that the point density of path is too high to further determine. Figure 13b shows the path obtained after process of Douglas-Peucker algorithm. The point density of path is far less than the original but keep the original complete path.

Fig. 12 The uncertainly coordinates of learner



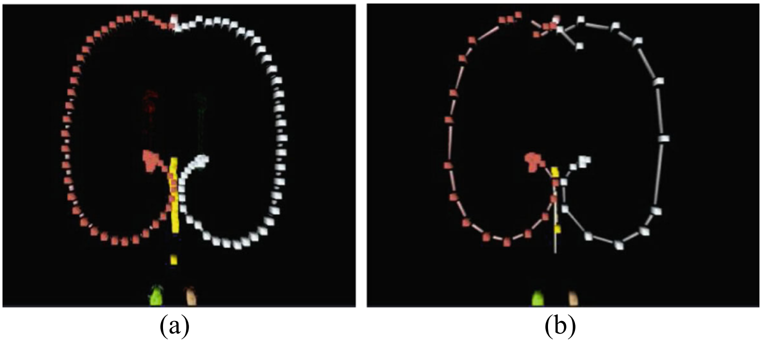


Fig. 13 **a** Original path from the Kinect capture. **b** Processed path after using Douglas-Peucker algorithm

3.1.4 Generating the guide path

Figure 14 shows the simplified continuous movement of guide path where divided into 4 fragments. Taking Fig. 14 as example, the movement of guide path has been divided into 4 fragments, each fragment composed of several continuous simplified points. During the guiding processing, the previous fragment would be disappeared when the next fragment appeared in screen. The fragments displayed in screen one by one to guide the learner to complete the correct movement.

In ideal situation, every point resolved by the same continuous movement should be closed in its original position. Under the premise of this assumption, each point will be made the relative time difference and the coordinate difference. As shown in Fig. 13, there are 14 relative time differences, T_{1-0}、 T_{2-0}…T_{13-0}、 T_{14-0} obtained; and there are 14 coordinate differences, P_{1-0}、 P_{2-0}…P_{13-0}、 P_{14-0} obtained. In next section, the process of comparing could be started when the movement is started to find the difference of relative time and coordinate. It is a reference to judge whether movement of learner is correct or not.

3.2 Part B. Comparison

Figure 15 shows the flow chart of comparison. Before examining the learner's movement, the learner has to be confirmed whether or not completes all movement at the correct time. Firstly, the

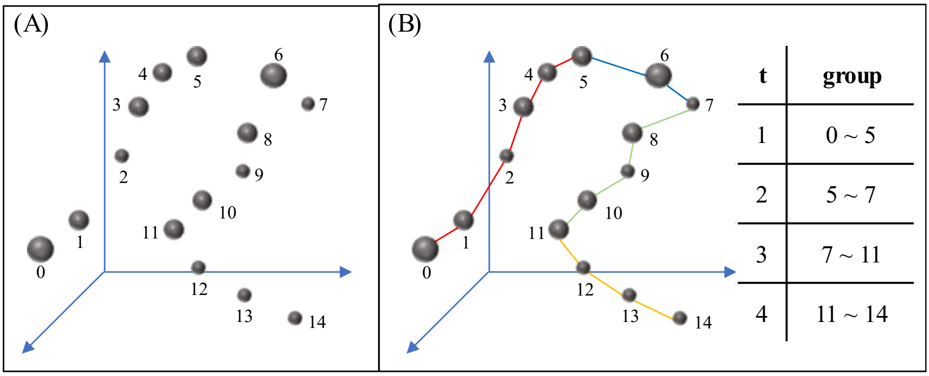


Fig. 14 Generating the fragments of guide path

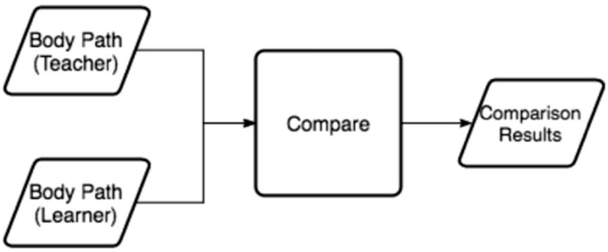


Fig. 15 Comparison flowchart

started coordinate of learner should be confirmed. The detection is started when learner starts to move and records whole path from starting to ending point. The system would compare the learner's movement of elapsed time with tutor's to decide the result whether follows up the tutor's or not.

3.2.1 Comparing

The learner's movement, follows up the guide path, is compared with the guide path. The continuous movement is divided into several fragments to be as comparison unit, the fragment is as well as shown in Fig. 14. There is a threshold has to be setup for the calculation of "correct rate" of learner's movement, as shown in Fig. 16. The correct rate is determined from the Eq. 3. It is a reference for a learner to know the correct rate to improve one's movement.

$$correct\ rate = 1 - \frac{distance}{threshold} \tag{3}$$

3.3 Part C. Interactive display

Figure 17 shows the interactive display combines of virtual tutor guide path and comparison result. The interactive display is able to provide the suggestions for learner to improve immediately. It includes mainly two parts:

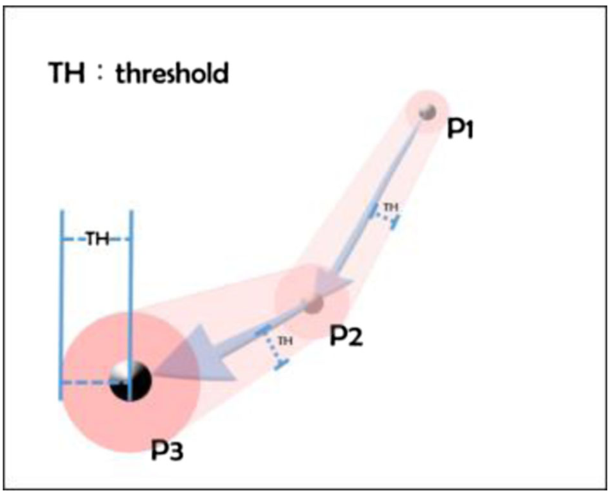


Fig. 16 The threshold for calculation of correct rate

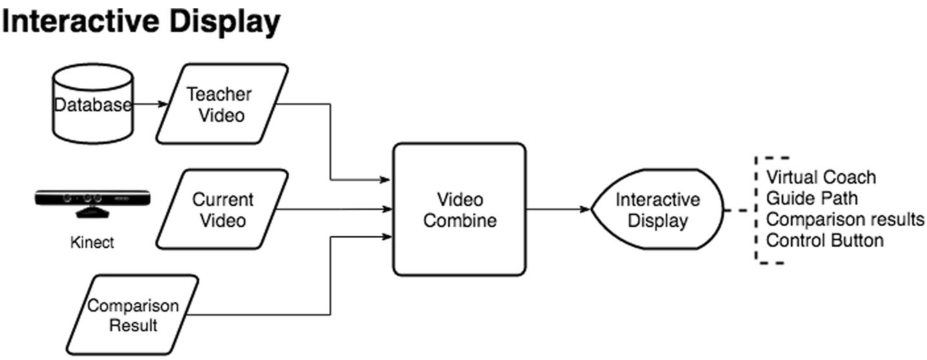


Fig. 17 Interactive display

i. The guide path of virtual tutor and learner movement
ii. The comparison results from the guide path and learner path

Learner obtains the information from the interactive display to modify one's movement.

4 Experimental Results

Figure 18 shows the demo of experiment in this system. The right side is the virtual tutor's movement; the left side is the learner's movement. It is found the guide path appeared on both virtual tutor and learner where guides the learner to move. The fragment of guide path disappeared when next fragment appeared in screen. The learner follows up the guide path fragment by fragment until the tutor's movement is completed.

Figure 19 shows two kinds of learner's movement in screen, comparison result and analysis. The first draw in both left and right side shows the learner follows up the path of virtual tutor's movement; the second draw shows the comparison of path between the guide and learner; and the third draw shows the analysis of this movement. It is found that the deviation of learner's path in right side is larger than the left, as shown in the second draw. The third draw shows the final analysis which reveals the correct rate of learner.

Fig. 18 Test situation

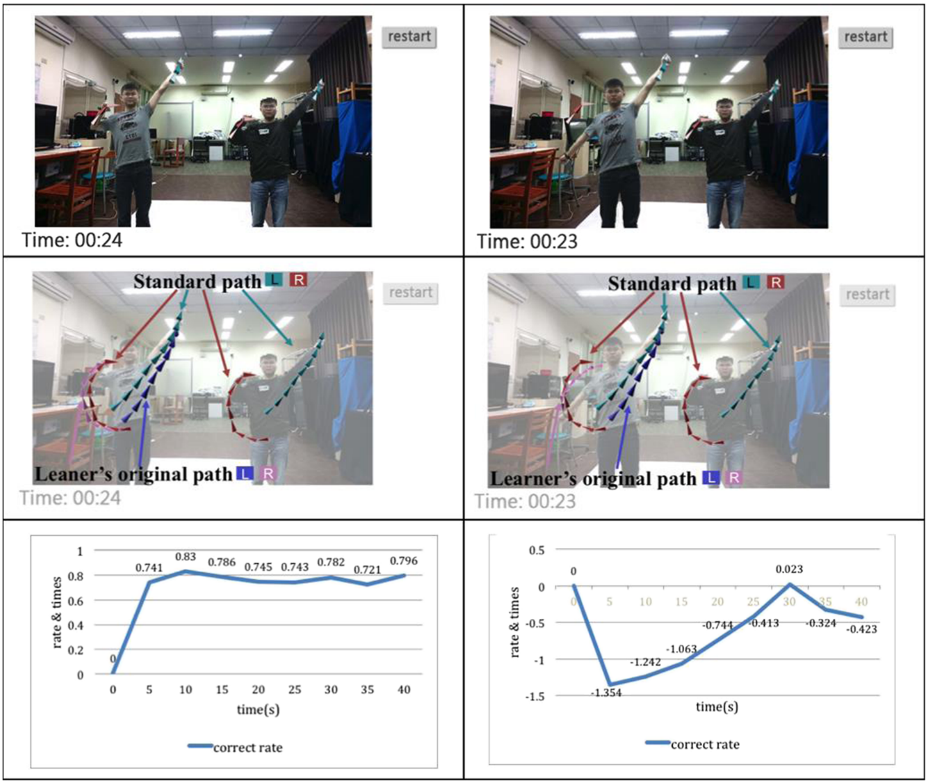


Fig. 19 Screen in movement, comparison and analysis

The virtual tutor learning system in this study achieves the purpose, the learner obtains the immediately feedback to learn the correct movement from virtual tutor without the constraint of space and time and with low cost by using this system.

There are some points should be further improved, discussed as following:

i. Backgrounds removing

The background removing is based on using the depth information, it would ignore the edges of the image. Taking the picture in Fig. 18 as example, the virtual tutor's hair and hand is incomplete because of the Kinect can't get the depth information from those area. If the RGB camera is added to use the function of block expansion or/and edge detection to obtain the integral area, the default would be made up.

ii. Guide path

The color guide path should be more obvious. The obvious color guide would provide more convenience for the learner to prompt one's correct rate in movement.

iii. Enhancing the tolerability of mistake make

There is a method to enhance the tolerability of mistake make. The movement could be acted by tutor repeatedly for many times and be recorded simultaneously. The generated guide path from each movement is collected and summarized as guide path of virtual tutor. This method would provide a higher enhancing tolerability of mistake make for learner.

5 Conclusions

This system provides an available design for a learner to obtain the movement train in three-dimensional space. In this experiment, we tested the system for the recognition of gesture segmentation of movement training. We intend to move beyond the traditional face-to-face and online training with low cost but without any constraint of space and time. The governing those movement eLearning training is developed from feedback immediately on combining capturing the whole skeleton correctly (video process, Kinect sensor), image treated (comparison) and interactive display. We modify the image obtained from Kinect sensor makes the database of eLearning is useful for learner of movement. This system of virtual tutor corrects the learner's movement with immediately feedback and analysis learner's movement in-situ. It is available to improve the current traditional face-to-face teaching and other eLearning system with low cost and high efficiently. We hope some of the limitation of our present work could be overcome in future work.

References

1. Alexiadis DS, Kelly P, Daras P, O'Connor NE, Boubekeur T, Moussa MB (2011) Evaluating a dancer's performance using kinect-based skeleton tracking. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 659–662
2. Bradley D, Roth G (2007) Adaptive thresholding using the integral image. J Graph Tools 12(2):13–21
3. Douglas D, Peucker T (1973) Algorithms for the reduction of the number of points required to represent a line or its caricature. Can Cartographer 10(2):112–122
4. Mita T, Kaneko T, Hori O (2005) Joint Haar-like features for face detection. In Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2 (ICCV '05), Vol. 2. IEEE Computer Society, Washington, DC, USA, 1619–1626
5. Moore JL, Dickson-Deane C, Galyen K (2011) e-learning, online learning, and distance learning environments: are they the same? Internet High Educ 14(2):129–135
6. Muneesawang P, Khan N, Kyan M (2015) A machine intelligence approach to virtual ballet training. IEEE Multimed 22(4):80–92
7. Oikonomidis I, Kyriazis N, Argyros A (2011) Efficient model-based 3-D tracking of hand articulations using Kinect. Proc. Brit. Mach. Vision Conf., BMVA Press, pp 101.1–101.11. https://doi.org/10.5244/C.25.101
8. Sehra S, Maghu S, Bhardawaj A (2014) Comparative analysis of E-learning and distance learning techniques. Int J Inf Commun Technol 4(8):823–828
9. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A (2011) Real-time human pose recognition in parts from single depth images. In Proceedings of the 2011 I.E. Conference on Computer Vision and Pattern Recognition (CVPR '11). IEEE Computer Society, Washington, DC, USA, 1297–1304
10. Vrettaros J, Tagoulis A, Giannopoulou N, Drigas A (2012) Case study in using Web 2.0 tools by Greek educators. Int J Soc Humanist Comput 1(4):363–374
11. Wilson PI, Fernandez J (2006) Facial feature detection using Haar classifiers. J Comput Sci Coll 21:127–133

Hsin-Hung Chiang received his M.S. degree in institute of computer science and information engineering from National I-Lan University, Taiwan, R.O.C. in 2009. He is currently pursuing a Ph.D degree in department of electrical engineering at National Dong Hwa University, Taiwan, R.O.C. His research interests include Image Process, Multimedia System and Peer-to-Peer Network.

Wei-Ming Chen is a professor of the Institute of Computer Science & Information Engineering, National Ilan University, where he also serves as director. He holds MS and PhD degrees from National Chung Cheng University. His research major is on multimedia communications and image processing.

Han-Chieh Chao (SM'04) received his M.S. and Ph.D. degrees in Electrical Engineering from Purdue University, West Lafayette, Indiana, in 1989 and 1993, respectively. He is currently a professor with the Department of Electrical Engineering, National Dong Hwa University, where he also serves as president. He is also with the Department of Computer Science and Information Engineering and the Department of Electronic Engineering, National Ilan University, Taiwan; College of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, China, and Fujian University of Technology, Fuzhou, China. He serves as the Editor-in-Chief for the Institution of Engineering and Technology Networks, the Journal of Internet Technology, the International Journal of Internet Protocol Technology, and the International Journal of Ad Hoc and Ubiquitous Computing. He is a Fellow of IET (IEE) and a Chartered Fellow of the British Computer Society.

De-Li Tsai received his M.S. degree in institute of computer science and information engineering from National I-Lan University, Taiwan, R.O.C. in 2017. His research interests include Image Process and Multimedia Application.