International Journal of Engineering & Technology, 7 (4.35) (2018) 294-298

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

AR Oriented Pose Matching Mechanism from Motion Capture Data

Javid Iqbal[1]*, Manjit Singh Sidhu[2], Mutahir Bin Mohamed Ariff [3]

[1,,2,3] Department of Graphics and Multimedia, College of Computer Science and IT Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Malaysia

**Corresponding author E-mail: manjavid@gmail.com*

Abstract

Pose matching and skeletal mapping method are an integral part of Augmented Reality (AR) based learning technology. In this paper a mechanism for pose matching is presented based on extraction of skeletal data from the dance trainer's physical movements in the form of color defined images snapped by Kinect, where each pose is modelled by a sequence of key movements and continues data frames. In order to extract the exact matched pose, the frame sequence is divided into pose feature frame and skeletal data frame by the use of pose matching dance training movement recognition algorithm (PMDTMR). This proposed algorithm is compared with other published methods in terms of frame level accuracy and learning time of dance session. The experimental results show that the proposed algorithm outperforms the state of art techniques for successful identification and recognition of matched pose between the dance trainer and the expert of the pre-recorded video through the Kinect sensor.

Keywords: *Augmented Reality; Dance training; Human action; Motion capture and Kinect sensor.*

1. Introduction

The recognition of human action is considered to be one of the challenging and crucial tasks in many application fields and research contributions such as interactions among robot and human, video surveillance, military and defense application and so on. With the advent of technology and exposure to internet huge progress has been correlated towards daily activity recognition and movement training. Mastering novel movements, posture and human motion is one of the essential factors of many physical activities such as martial arts, sports physical exercises and dancing. The process of expertising these motor skills can be a challenging task requiring hours of training and repetitive practice. Moreover, the ability of affordable Kinect sensor and technology advancements has enabled movement training to occur at home. It can be envisaged that the process of reacting and observing human gestures will also become an important skill for human made robots in addition to technology-based dance learning.

The color defined data provides the human movement postures which can be apprehended by Kinect camera sensor. The motion analysis process that is carried out by Kinect sensor tracks the skeleton of the person performing physical movements. The 3D image of the trainer is also detected and comprehended by Kinect sensor using visible and IR cameras. Moreover, the human motion recognition can be modelled as a continues evolution of skeletal joints. The state of art pose matching and human action recognition techniques have taken into account only segmentation and unified action classification. Thereby providing means to identification of a single person performing a single movement action. Therefore, the existing pose recognition and matching algorithm may face issues such as one RGB data sequence containing several clumsy human gesture and inability to identify starting and ending of each movement frame.

2. Background Study

The amalgam combinations of traditional fine art forms and advanced techniques has made the researchers in the last decade to opt for dance training solutions based on AR. The Cha learn gesture captures data which is dependent on the user, utilizes the miniature vocabulary and single shot training based on Kinect camera was proposed in [1]. Authors of [2] and [3] had presented method of visualization based on the acceleration, speed and velocity of the motion of the dance teacher for clearer and effortless understanding of the students. Researchers of [4] examined a novel framework YouMove that enables student to record and learn physical development successions. Figure 1 shows the system design of YouMove.

The tracking and improvised gesture recognition system is termed as Action Graph (AG) was proposed by [5] which has the ability of catching approaching motions in an unsupervised way and empowers mapping between input motions to wanted rendering functionalities. Researchers in [6] have proposed a novel optimizing solution for efficient and effective motion detection in Kinect-based entertaining environment. Authors of [7] proposed a novel procedure for ballet e-learning learners in the remote areas by put on the benefit of type-1fuzzy set.

The authors in [8] had proposed a Kinect based training system for slow movements where the method of assessment is to evaluate the similarity of posture and timing between the trainer and user. The approach discussed in [9] investigates the Vietnamese folk dances by applying dancer's body part movement analysis through representation of an initial experimental ontology. The Thai dance training tool system is designed and prototyped using Kinect motion sensor in [10]. The significant research contribution in [11] describes the classification of Indian traditional dance forms from videos, extraction of these representations through Deep Convolution Neural Networks (DCNN) and optical flow followed by training over a multiclass linear Support Vector Machine (SVM). Different dance styles that can be recognized is described in Figure 2. The researchers in [12] have explored an application of Kalman filters for dance visualization in real-time within the unity3d conceptual framework. In [13], a Thai dance training game-based model is proposed and developed which serves as a guideline framework to promote learning motivation using Kinect-based training system.

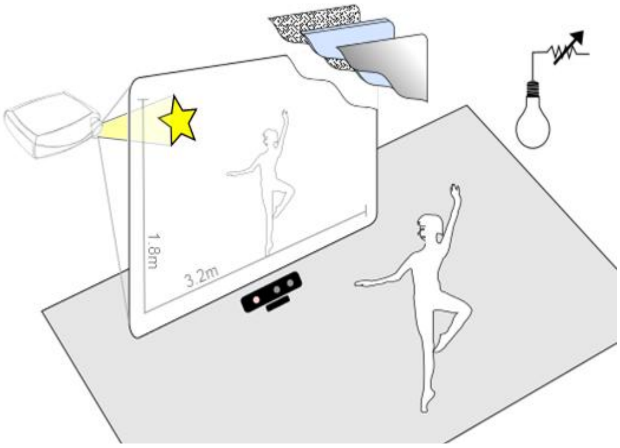

Fig. 1: System design YouMove [4]

Fig. 2: Different types of dance form [11]

3. Methodology

We designed, developed and implemented a Natural User Interface (NUI) based dance training system, the theoretical framework of which has been depicted in Figure 3. The system has been implemented using a Kinect camera for four different dance style- Bharatanatyam, Western, Bollywood and Joget (Zapin). The Kinect tracks the movements and provides scores for the trainee based on the skeletal mapping and comparison to the choreographer's movements in the video. The performance of the system was tested and validated with one on one traditional dance learning method inclusive of teachers with dance expertise. This paper describes one of the mechanisms of the proposed research which is the pose matching mechanism.

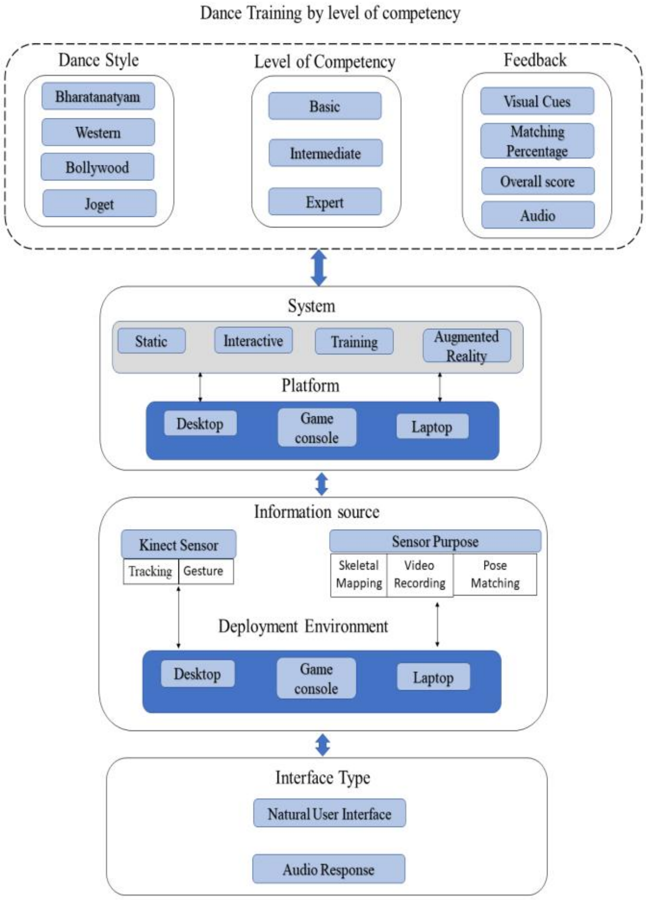

Fig. 3: Theoretical framework

3.1 Dance Learning System Using AR

3.1.1 Advantages of AR based Learning

There are vast number of AR advantages in education that are user friendly, interactive and attention seeking for dance training and learning by [14]. The training based on AR technology includes the features such as higher responsiveness, user friendly, understandability, cost effectiveness, interactive, highly adaptive. The features of AR based learning include robust, cheaper, user friendly, animated, timely, easily understandable, responsive, seamless integration with other media, leading to higher reliability, portability, efficiency and interesting solutions as discussed by researchers of [15]. The advantages and the features of AR is depicted in Figure 4.

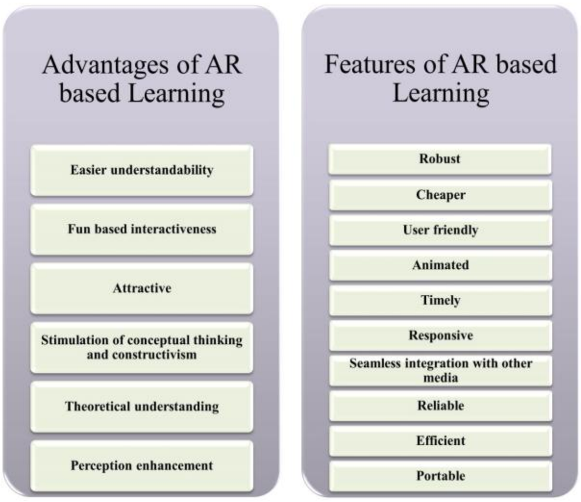

Fig. 4: Advantages of AR based learning [14]

3.1.2 AR Based Environments and Implementation for Dance

The researchers of [4] have described a unique framework YouMove that enables users to record and learn physical movements. The Kinect-based framework is intended to be straightforward, enabling anybody to make and offer preparing content, the training system uses recorded data to train the user using a large-scale AR mirror. The authors have also presented a user study in which YouMove was shown to improve training and short-term retention by a factor of 2 compared to a conventional video demonstration. Digital Bharatanatyam interaction framework was developed by researcher in [16].
A Hierarchical model and bottom up process flow for implementing Bharatanatyam dance was also explained. The authors of [7] proposed a novel strategy for ballet e-learning for novices in remote areas. This method is focused on training dataset of thirty-four different ballet postures, obtained from images of trained dancers. The researchers also claim that the quality performance of the proposed algorithm is substantiated by the reported simulation results, with an accuracy of 91.23%.
Evaluation is a specific feature obtained by getting the training data set T_s for a specific posture *d*, the mean $\bar{g}_{m,n}{}^{\rho}$ and the standard deviation $P_{m,n}{}^{\rho}$ are computed across K image from equation 1 and 2.

$$\bar{g}^{P}_{m,n} = \frac{1}{K}\sum_{k=1}^{k} \alpha^{p,k}_{m,n} \tag{1}$$

$$P^{P}_{m,n} = \sqrt{\left(\frac{1}{K}\sum_{k=1}^{k} \alpha^{p,k}_{m,n} - \tilde{\alpha}^{p}_{m,n}\right)} \tag{2}$$

In this way, the local occupancy information of this bin is

$$P_{xyz=}\,\delta\left(\sum_{q\in bin_{xyz}} D_q\right) \tag{3}$$

where D_q = 1 if the point cloud has a point in the location q and D_q = 0 otherwise. δ(.) is a sigmoid normalization factor. The Local occupancy pattern (LOP) feature of a joint i is a vector consisting of the feature P_{xyz} of all the bins in the spatial grid around the joint, denoted by Pi.

3.2. Pose Matching Mechanism

In the proposed algorithm firstly, a class method is declared for identification and recognition of specific pose by the dance trainer. Secondly, the reference joints and the joints of interest extracted from the individual frame of RGB-D image and skeletal data are identified. The relative and normalised orient (RNO) for each pose is calculated based on the following formula,
Let,

$$M_i = \left(m^i_x, m^i_y, m^i_z\right)$$

Denotes the position determined by the joints in 3D world system of coordinates, where the RNO of 'i' joint relative to 'j' joint can be mathematically calculated as,

$$G_{NRO} = \left(\frac{M_i - M_j}{||M_i - M_j||}\right)$$

Where ||..|| signifies the distance of Euclidean points. It can be derived from the above explanation and mathematical formulas that RNO is resistive and has seldom effect with respect to the user's distance, height and limb length from/ towards the Kinect camera sensor.
The condition for matching the pose of left hand and right hand above hip by the dance instructor is checked along with a declaration of a Boolean variable is followed as subsequent step. The next condition for pose matching of leg is checked and the distance of the leg L_D from the Kinect sensor and the leg distance threshold L_{DT} is computed using the following equation,

$$L_D = Abs|\,X_{KR} - X_{KL}|$$
$$L_{DT} = S_H * 0.7f$$
$$L_E = L_D > L_{DT}$$

Where, X_{KR} and X_{KL} are the X coordinates of right ankle and left ankle extracted by the Kinect and S_H refers to length between spine to head and 0.7f is the coefficient of threshold float value. L_E denotes the Boolean variable declared for the outcome of leg movement extraction and segmentation fragmentation.
Lastly the recording of the dance learning session is started and the total learning time for each pose matched is computed followed by motivating the trainer with interactive remarks for each pose that is matched from the RGB-D image of the video frame. The outcome of the pose matching algorithm and the mechanism is real-time are shown is Figure 5 and Figure 6.

3.2.1 Algorithm: Pose Matching Mechanism

```
Initial pose: Arms are wide

Step1: Declare a Simple class DancePose's specific pose recognition

Step 2: Create a class to implement DancePose.RecognizePose()

Step 3: Find the joints of interest and reference joints using
KinectInterop.JointData hipL = bd.joint[(int)KinectInterop.JointType.HipLeft];
KinectInterop.JointData hipR = bd.joint[(int)KinectInterop.JointType.HipRight];

Step 4:  Identify the joints in motion using
KinectInterop.JointData handL = bd.joint[(int)KinectInterop.JointType.HandLeft];
KinectInterop.JointData handR = bd.joint[(int)KinectInterop.JointType.HandRight];

Step 5:
        If
            bool leftHandAboveHip = handL.kinectPos.y > hipL.kinectPos.y;
        else
            bool rightHandAboveHip = handR.kinectPos.y > hipR.kinectPos.y;
End

            Check if the the legs are apart enough using the condition
        If
            float legDistance = Mathf.Abs(ankleR.kinectPos.x - ankleL.kinectPos.x);
        else if
            float legDistanceThreshold = spineToHeadLength * 0.7f;
        else
            bool legsExtended = legDistance > legDistanceThreshold;
End

Step 6: Set visualization modes based on if the left leg moved properly and arms are up

Step 7: Overwrite debug values set in base.RecognizePose()

Step 8: For each timeframe i
for (int i = 0; i < 6; i++)
loopStart = i * 12 * beat + intro;

Step 9:
Start  RecordingBehavior startRec = RecordingBehavior.Start;
if(i != 0)
startRec = RecordingBehavior.None;
        End
End

Step 10: For each pose matched Compute
targetPosesTimeline.Enqueue(new RightSideStepArmsLifted(loopStart + beat * 0, beat, startRec));

Repeat the same from step 2 to step 10 for each pose from the frame.
```

Fig. 5: Pose matching algorithm

Fig. 6: Pose matching mechanism

4. Results and Discussion

The comparative analysis of the proposed algorithm is carried out in terms of performance metrics such as accuracy of the pose matched and learning time. The poses of the dance trainer are snapped by the sensors of the Kinect camera and only the skeletal extraction of colour defined images after applying the proposed algorithm are utilized in the experiments. The poses that are matched includes the basic beginner steps of western dance. The framerate is 30fps and data sequence length is about 60 secs. The total number of activity sequence is 90, the steps are performed by a single trainer in indoor environment in compliance to a pre-recorded video of a dance expert. The test data sequences are combined into one continues data sequence to provide the marking scheme for the poses matched after applying the proposed algorithm through Kinect sensor.

The frame level-based accuracy is utilized as criteria to evaluate the proposed PMDTMR algorithm and compare it with another existing algorithm. The poses that are matched is labelled framewise a priori and the result is obtained in frame level. Thus, the frame level accuracy is computed based on the ratio in between the number of correctly matched frames and the total number of frames in the data sequence. The number of matched poses for the western dance style is Wp=90, the segmentation threshold is set to 0.05 and the maximal length of head to hand is LP_{max} = 0.8f. The experiments are developed, implemented and executed on the system with Intel core i7, NVIDIA® GeForce® 940MX with 12 GB RAM and Windows 10 operating system. MATLAB 2017b is employed as a simulated means for the function of GMM that is build-in function of MATLAB. The overall efficiency of the proposed mechanism is improved through the utilization of MATLAB with C# language.

The comparative analysis as depicted in Figure 7 and 8 shows that the proposed pose matching algorithm has higher frame level accuracy for the higher probability of matched poses. Moreover, the learning time for a particular dance style is also considered to be lesser when compared to the state of art algorithm using the Kinect sensor.

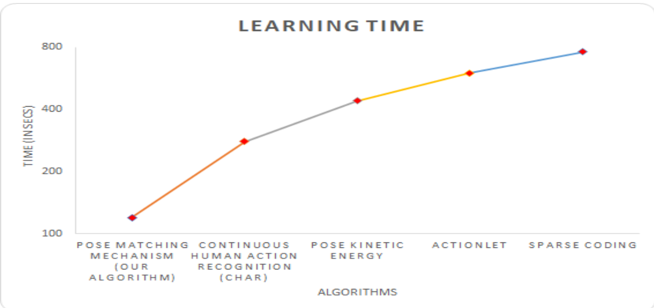

Fig. 7: Comparative analysis learning time

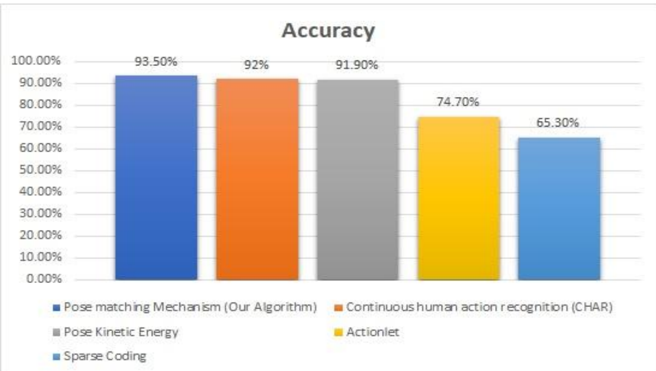

Fig. 8: Comparative analysis accuracy

Table 1: Accuracy

Method	Accuracy
Pose matching Mechanism (Our Algorithm)	93.50%
Continuous human action recognition (CHAR) [17]	92%
Kinetic Energy Pose [18]	91.90%
Actionlet [5]	74.70%
Coding Sparse [19]	65.30%

Table 2: Learning time

Method	Learning time in sec
Pose matching Mechanism (Our Algorithm)	120
Continuous human action recognition (CHAR)	280
Kinetic Energy Pose	440
Actionlet	600
Coding Sparse	760

5. Conclusion

In this paper a pose matching dance trainee movement recognition algorithm is proposed. This algorithm extracts the skeletal data of the whole human body and matches the pose with the dance expert of the pre-recorded video. The skeletal data extracted from the RGB-D images of the Kinect sensor performs the matching and the reorganization process on the segmented frames obtained by the proposed algorithm. The effectiveness and the efficiency of the proposed method is tested and evaluated in terms of frame level accuracy and the cumulative learning time of the dance steps. The authors are currently working on other modules of the research which are skeletal mapping and gesture recognition according to Figure 3, by integrated utilization of skeletal data, data sequences/ frames of matched poses, RGB-D images and trainer- Kinect NUI interaction features.

Acknowledgement

We would like to thank iRMC for proving the research fund for this project under UNITEN Internal Grant (**UNIIG**)

References

[1] I. Guyon, V. Athitsos, P. Jangyodsuk, H. J. Escalante, and B. Hamner, “Results and Analysis of the ChaLearn Gesture Challenge 2012,” Springer,WDIA 2012, LNCS, pp. 186–204, 2012.

[2] M. Eichner and V. Ferrari, “Human pose co-estimation and applications,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2282–2288, 2012.

[3] I. Kuramoto, Y. Nishimura, K. Yamamoto, Y. Shibuya, and Y. Tsujino, “Visualizing Velocity and Acceleration on Augmented Practice Mirror Self-Learning Support System of Physical Motion,” 2013 Second IIAI Int. Conf. Adv. Appl. Informatics, pp. 365–368, 2013.

[4] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, “YouMove: enhancing movement training with an augmented reality mirror,” Proc. 26th Annu. ACM Symp. User interface Softw. Technol. - UIST ’13, pp. 311–320, 2013.

[5] J. Wang, Z. Liu, and Y. Wu, “Learning Actionlet Ensemble for 3D Human Action Recognition,” SpringerBriefs Comput. Sci., vol. 36, no. 9783319045603, pp. 11–40, 2014.

[6] X. Chang, Z. Ma, M. Lin, Y. Yang, and A. G. Hauptmann, “Feature Interaction Augmented Sparse Learning for Fast Kinect Motion Detection,” IEEE Trans. Image Process., vol. 26, no. 8, pp. 3911–3920, 2017.

[7] S. Saha, P. Rakshit, and A. Konar, “Ballet E-learning using fuzzy set induced posture recognition by piece-wise linear approximation of connected components,” Appl. Soft Comput. J., vol. 65, pp. 554–576, 2018.

[8] I. V. Masala and A. Angdresey, “The Real Time Training System with Kinect: Trainer Approach,” IEEE, 2017 Int. Conf. Soft Comput. Intell. Syst. Inf. Technol., 2017.

[9] M. T. Chau, “A Labanotation based ontology for representing Vietnamese folk dances,” 2018 Int. Conf. Digit. Arts, Media Technol., pp. 75–80, 2018.

[10] Y. Tongpaeng, A. Technology, and T. Dance, “Evaluating Real-Time Thai Dance Using Thai Dance Training Tool,” 2018 Int. Conf. Digit. Arts, Media Technol., pp. 185–189, 2018.

[11] A. Bisht, R. Bora, G. Saini, P. Shukla, and A. Bisht, “Indian Dance Form Recognition from Videos,” 2017 13th Int. Conf. Signal-Image Technol. Internet-Based Syst., pp. 123–128, 2017.

[12] K. Abramova, A. Corradini, and J. Nordentoft, “Real-time motion tracking for dance visualization using Kalman filters,” Ieeexplore.Ieee.Org, pp. 343–347, 2018.

[13] S. Chernbumroong and K. Tabia, "Thai Dance Training Game-Based Model," 2018 Int. Conf. Digit. Arts, Media Technol., pp. 194–197, 2018.
[14] J. Iqbal and M. S. Sidhu, "A review on making things see: Augmented reality for futuristic virtual educator," Cogent Educ., vol. 4, no. 1, pp. 1–14, 2017.
[15] M. Lee, K. Lee, and J. Park, "Music similarity-based approach to generating dance motion sequence," Multimed. Tools Appl., vol. 62, no. 3, pp. 895–912, 2012.
[16] R. Majumdar, "Framework for teaching Bharatanatyam through Digital Medium," IEEE Fourth Int. Conf. Technol. Educ. Framew., pp. 241–242, 2012.
[17] G. Zhu, L. Zhang, P. Shen, and J. Song, "An Online Continuous Human Action Recognition Algorithm Based on the Kinect Sensor," Sensors, vol. 16, no. 2, p. 161, 2016.
[18] J. Shan and S. Akella, "3D human action segmentation and recognition using pose kinetic energy," IEEE Int. Work. Adv. Robot. its Soc. Impacts, pp. 69–75, 2014.
[19] B. Ni, P. Moulin, and S. Yan, "Order-Preserving Sparse Coding for Sequence Classification," Proc. Comput. Vision–ECCV, Firenze, Italy, Oct. 2012, pp. 173–187, 2012.