Dance Motion Detection Algorithm Based on Computer Vision

YanWang[1], Zhiguo Wu[2]
EIT Data Science and Communication College, Zhejiang Yuexiu University, Shaoxing 312000, Zhejiang, China[1]
School of Art, Zhejiang Yuexiu University, Shaoxing 312000, Zhejiang, China[2]

***Abstract*—Human posture recognition is an essential link in the development of human-computer interaction. Currently, the existing dance movement training methods often require students to constantly watch videos or find a tutor to correct them during practice to achieve good results, which not only takes a lot of time and energy but also creates some difficulties and challenges for students. The research goal of this paper was to use computer recognition technology to detect dance movements and identify body postures. This paper develops a Kinect dance auxiliary training system based on the body skeleton tracking technology of the Kinect 3D sensor, combined with auxiliary dance training. This article not only introduced a fixed axis-based expression method for joint angles to improve the stability of joint angles but also improved the body position detection algorithm using the angle of joint spots to realize the accurate recognition of human body posture. In the experiment, the trainee's arm was raised to the highest position, which could not meet the requirements, and the trainer's wrist should be raised by another 200 mm. Moreover, retracting the hand was too fast, which did not meet the standard action. The test results showed that the system could effectively improve the dance movements of the students.**

Keywords*—*Dance motion detection; computer vision; human posture recognition; Kinect 3D sensor

I. INTRODUCTION

Dance is a captivating and culturally significant form of expression that has gained popularity not only in the realm of entertainment but also in educational and daily life contexts. Achieving excellence in dance requires rigorous adherence to standards, often necessitating the guidance of specialized instructors. However, the subjective nature of human judgment and the lack of standardized guidance in dance practice present challenges for learners. Furthermore, relying solely on expert feedback can limit the flexibility and accessibility of dance training. This underscores the increasing importance of leveraging intelligent machinery to provide precise regulation and supervision of dance movements, particularly in scenarios where human guidance is unavailable.

In the backdrop of a rapidly evolving society, the field of dance motion detection has seen a surge in research interest. Recognizing the crucial role of subtle, momentary cues in synchronized movements, recent studies have delved into exploring multifaceted dimensions of dance. For instance, Skoe et al. [1] have investigated the nuanced objectives of worldly dance, shedding light on the intricate connection between dance cognition and auditory perception. Thornquist [2] has explored the potential of dance exercise therapy as a means to reduce individuals' reliance on self-expressive substances. Additionally, research methodologies have emerged to assess the efficacy of diverse dance modalities. Reddy and Pereira [3] have proposed a compelling link between performers' and audiences' inner experiences, suggesting the possibility of generating non-local interactions through dance. Sun et al. [4] have introduced a pioneering dance self-learning framework grounded in Laban Movement Analysis (LMA) principles, enabling trainees to autonomously analyze and refine their. However, the algorithm for dance motion detection can be described in combination with sensors and human bones. Shiratori et al. [5] investigated motion structure detection by use of musical information and classified them as primitive motions. Chang et al. [6] employed the Feature Interaction Augmented Sparse Learning algorithm for motion detection based on three public Kinect-based entertainment.

Computer vision has become increasingly important and effective in recent years due to its wide application in many fields. Barbu et al. [7] proposed a new efficient learning scheme that tightens the sparsity constraint by gradually removing variables based on criteria and schedules. His experiments on real and synthetic data showed that the proposed method outperformed other state-of-the-art methods in regression, classification, and ranking while being computationally efficient and scalable and providing a foundation for dance motion detection. Cha et al. [8] proposed a damage detection method by integrating nonlinear recursive filters and non-contact computer vision-based algorithms to measure structural dynamic responses. His experimental results showed that the prediction of stiffness and damping characteristics was reasonable and accurate compared with dynamic analysis calculations, which gave a new idea for dance motion detection. Khan et al. [9] described and discussed case studies on applying Convolutional Neural Networks (CNN) in computer vision, including image classification, object detection, and semantic segmentation. The case study provides more contrastive content for dance motion detection algorithms. Fang et al.[10] developed an automatic computer vision-based method that uses two CNN models to determine whether a worker is wearing a seat belt while working at heights, which plays a good role in promoting a dance motion detection algorithm. However, these algorithms do not consider the complex factors affecting reality, so the algorithm's performance needs to be improved. In another study, Cha et al. [11] proposed a CNN architecture for diagnosing the weaknesses of concrete cracks with no fault detection extraction. Roberts et al. [12] employed the CNN for the classification of cranes to monitor the safety dangers with utilizing UAVs (Unmanned Aerial Vehicles). Ding et al. [13] recommended an integrated learning model that employs the CNN and long-short-term memory (LSTM) combination to examine the unsafe treatment of the workers. Lecun et al. [14] investigated a combination of CNN and the Mixed National Institute of Standards and Technology (MNIST) for handwritten number detection.

Aiming at the problem of the regulation and supervision of dance movements when performing independent dance exercises without human guidance, this paper used the Kinect method to detect dance movements. Through gesture recognition, key point feature processing, and action classification, a human pose estimation algorithm was established, and motion recognition was performed. The innovation of this paper is that a set of dance assistant training systems based on Kinect is constructed based on Kinect. The system is able to receive and identify trainers' dance steps automatically. Standard moves are not compared in the database and evaluated regarding joint point coordinates and angles, so corresponding operation comparison charts and guidance can be provided.

In the presented paper, a comprehensive methodology is presented in Section II. The experimental procedure to train the dance according to Kinect in conjunction with a discussion on the obtained data is presented in Section III. Eventually, a brief conclusion is brought out about the conducted study.

II. Methods of Human Pose Recognition

A. Computer Vision

Humans obtain information through hearing, vision, taste, touch, and other senses, among which vision is the most important source of information. Therefore, replacing human eyes with computers for observation has become a hot topic. The perceptron of a computer is generally composed of different image sensors and electronic images. It functions like the human brain, which collects, transmits, stores, and processes image information. Computer vision technology aims to enable computers to observe the outside world independently, recognize the specific information it contains, and adapt to the environment. It can observe and understand the world like a human instead of using the eyes to do specific work. In computer vision technology, acquiring and understanding static objects is relatively easy. However, for the human body, the activities of the body are more dynamic, so it is very meaningful to capture and analyze the action posture of the human body [15], [16].

Dance training is complex and multifaceted, while basic training cultivates basic skills. The purpose of the training is to exercise the strength of each part of the trainer's body and the extensibility of the joints so that the trainer can better control his body when performing actions such as jumping, rotating, and kicking, thereby achieving better results [17], [18]. Based on retaining the original dance form has become a difficult problem in the current dance teaching to achieve accurate and graceful movements.

Kinect was originally used in games and has been used in other ways, as shown in Fig. 1. Since Kinect has open source code, low prices, rich imaging, and powerful performance, many scholars have combined the Kinect platform to develop many different applications while developing somatosensory games [19].

Kinect can get profundity pictures of the human body through profundity distance. Kinect's imaging guideline depends on optical coding. Its functioning standard is displayed in Fig. 2. The laser shoots the laser light on the article, which the crooked medium dissipates. The laser produces a dot, which is projected on the item. Then, the spot is caught by the sensor. Because of the arbitrariness of the spot, it is important to utilize likelihood and factual techniques to break down the dot's differentiation, force dispersion, and development laws. The dissemination of laser spots is likewise arbitrary and has specific factual regulations [20], [21].

In order to solve the problem of frame rate, the theoretical frame rate of the Kinect sensor is 30 frames per second, and the depth map of each frame can reflect the true depth of a person. The depth image taken after a period can record the next person's real-time location [22]. Kinect can identify and track human bones. First, 25 key spots of the body are identified. Then, the body's skeletal structure is constructed and combined with depth information to form a three-dimensional human skeleton. The skeleton joint point model is shown in Fig. 3.

In order to identify the human body and track the bones of the target, it is first necessary to perform a deep scan on it to obtain the corresponding bone point information, which is then converted into a complete human body model. The color image is a two-dimensional image coordinate system (X, Y), and the depth image is a three-dimensional coordinate system. In the human skeleton coordinate system, (X, Y, and Z) are used as the coordinates. Bone tracking technology separates the background from the human body. Methods such as matrix transformation and machine learning are used to identify the key parts and three-dimensional coordinates of human skeletons combined with depth data, which provides convenience for future research [23]. Fig. 4 shows the skeleton tracking process of Kinect.

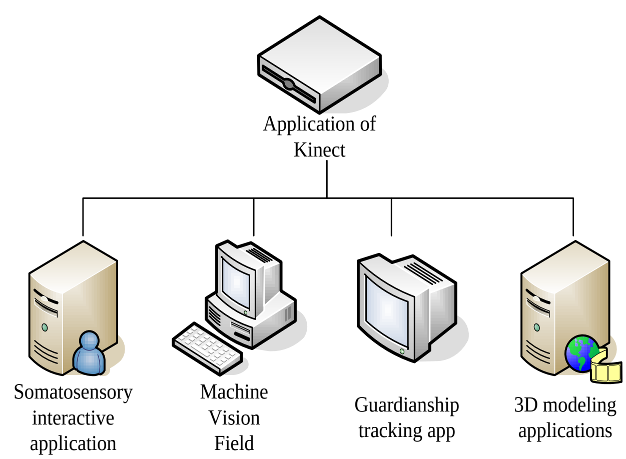


Fig. 1. Application of Kinect.

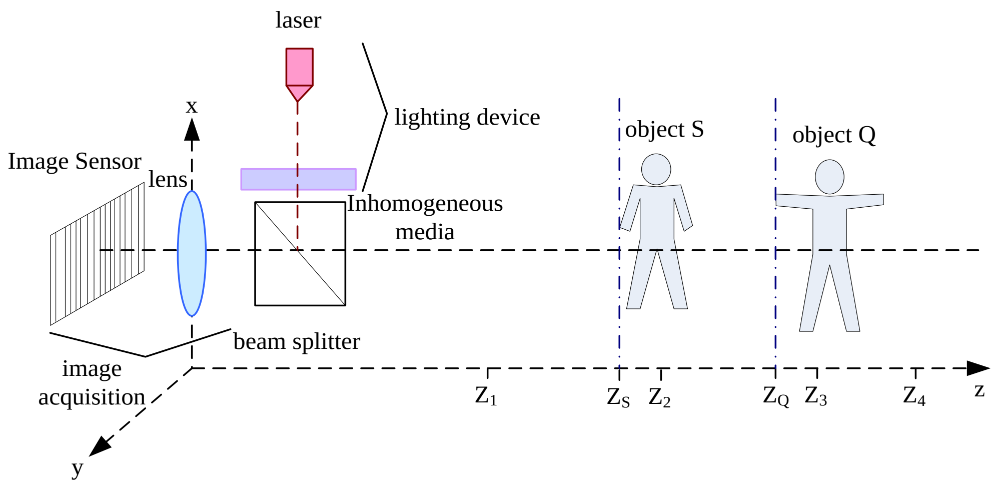


Fig. 2. Depth imaging system.

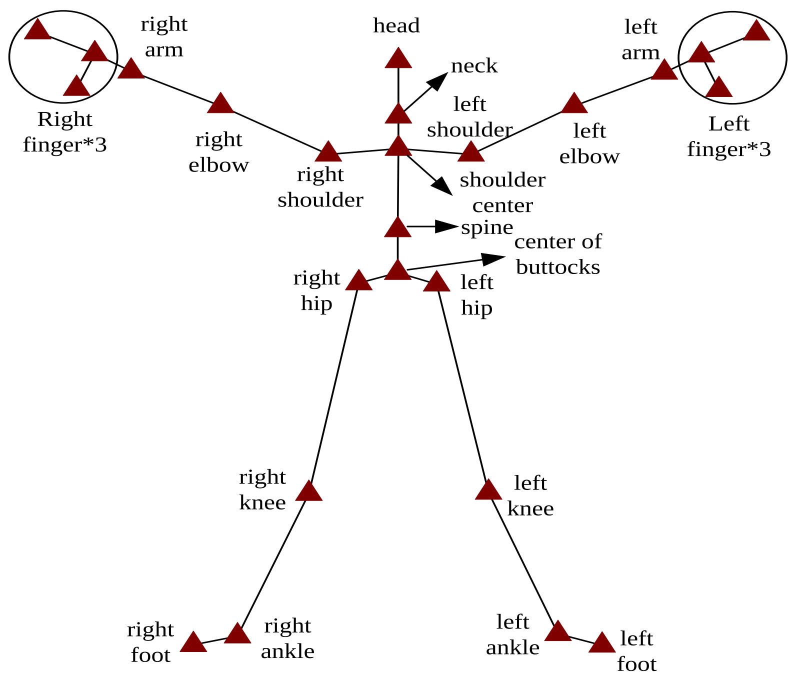


Fig. 3. Human skeleton joint spot figure identified by Kinect.

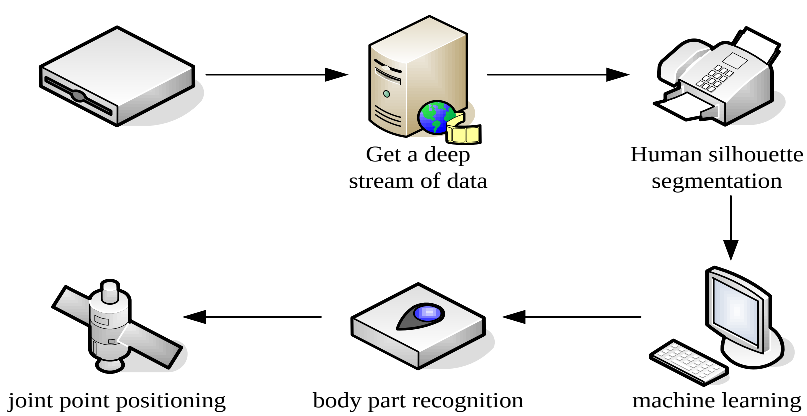


Fig. 4. Specific steps of Kinect bone tracking.

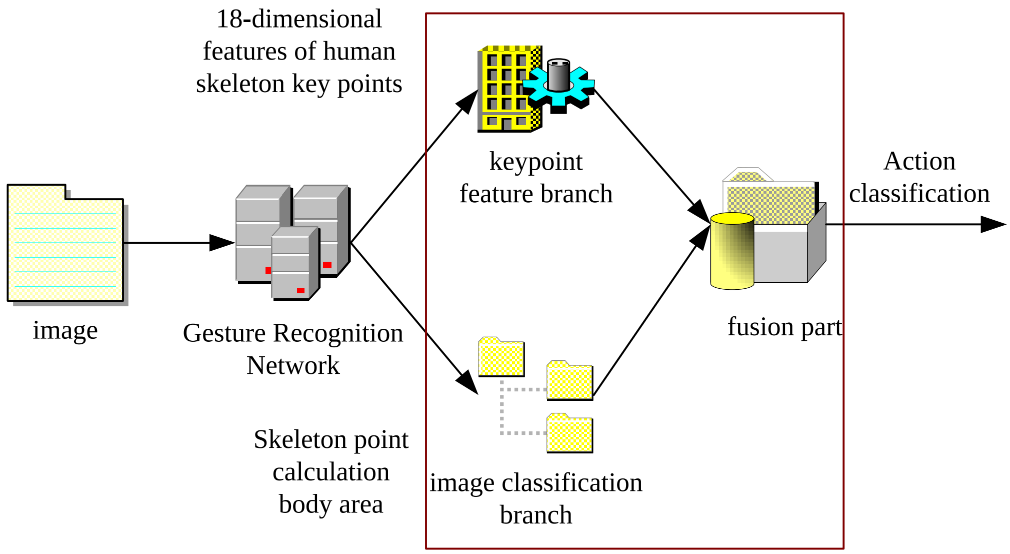


Fig. 5. Dance action detection based on gesture recognition.

The per-pixel information can be obtained through body part recognition inference, which defines a density estimate for the body part as follows:

$$f_v(\hat{c}) \propto \sum_{o=1}^{M} e_{ov}\, exp\left(-\left\|\frac{\hat{c}-\hat{c}_o}{q_v}\right\|^2\right) \quad (1)$$

Through Formula (1), the coordinate $\hat{c}$ in a three-dimensional space is obtained. M is the number of pixels and e_{ov} is the weight of the pixels. $\hat{c}_o$ Represents the projection of pixel $\hat{c}$. q_v represents the width of each component. e_{ov} is balanced between pixel inference and the probability of spatial regions.

$$e_{ov} = A(v|O, c_o) \cdot h_z(c_o)^2 \quad (2)$$

This approach improves joint prediction accuracy while guaranteeing a constant density estimate for depth.

This paper introduces a gesture-based action recognition method. First, the image is divided into a posé recognition network with a size of 368*368. The residual network recognizes the key spots of the body. Combining the feature classification of key points and image classification, the classification of the human body can be completed, and the classification of dance movements can be achieved. The specific process of dance action detection based on gesture recognition is shown in Fig. 5.

B. Establishment of the Human Skeleton Model

In attitude estimation, when using computer vision for attitude estimation, the first thing to solve is how to detect the human body, that is, the detection of moving objects. The recognition of moving objects is the premise of pose estimation and action recognition, and it is also the basis of skeleton tracking technology in pose estimation. The detection of moving objects is a method of using the flow information of light, or according to the difference between frames, to find the moving points of moving objects to realize the extraction of moving objects [24]. The algorithm is the first step in the human body's pose estimation and poses recognition. The moving objects are extracted to remove redundant background information, which is only analyzed for moving objects to achieve better results and accuracy.

Because the data acquisition in Kinect is 30 frames per second, the movement of the human skeleton can be regarded as an accelerated linear movement. L'is the number of frames, that is, discrete moments. Then, $x_o(l)$ can represent the position of the joint point o in the X-axis direction at time $l\Delta u$. $\dot{x}_o(l)$ Is the speed of the joint point o in the X-axis direction at time $l\Delta u$. If any point in the body is defined as o, then the Taylor expansion of the formula for the position and velocity of the articulation point on the x-axis is as follows:

$$x_o(l) = x_o(l-1) + \dot{x}_o(l-1)\Delta u + \frac{\Delta u^2}{2!}\ddot{x}_o(l-1) + \ldots \quad (3)$$

$$\dot{x}_o(l) = \dot{x}_o(l-1) + \ddot{x}_o(l-1)\Delta u + \frac{\Delta u^2}{2!}\dddot{x}_o(l-1) + \ldots \quad (4)$$

Similarly, in the Y-axis and Z-axis directions, the position and velocity can be expanded by Taylor. Then, according to Formula (3) and Formula (4), the mathematical model is filtered:

$$x_o(l+1) = SX_o(l) + E(l) \quad (5)$$

Among them:

$$x_o(l+1) = \begin{bmatrix} x_o(l+1) \\ \dot{x}_o(l+1) \\ y_o(l+1) \\ \dot{y}_o(l+1) \\ z_o(l+1) \\ \dot{z}_o(l+1) \end{bmatrix} \quad S = \begin{bmatrix} 1 & \Delta u & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta u & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta u \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

$$E(l) = \begin{bmatrix} e_{x_o}(l) & e_{\dot{x}_o}(l) & e_{y_o}(l) & e_{\dot{y}_o}(l) & e_{z_o}(l) & e_{\ddot{z}_o}(l) \end{bmatrix}^U \quad (7)$$

S is the state transfer matrix of the system. It estimates the movement of skeletal points through the $W(l)$ -diagonal matrix method and the process noise covariance matrix $E(l)$.

The following is the mathematical model for the observation matrix :

$$Z_o(l) = JX_o(l) + B(l) \tag{8}$$

Among them:

$$Z_o(l) = \begin{bmatrix} x_o^n(l) \\ y_o^n(l) \\ z_o^n(l) \end{bmatrix} \cdot J \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, B(l) \begin{bmatrix} b_x(l) \\ b_y(l) \\ b_z(l) \end{bmatrix} \tag{9}$$

The observed variable is $Z \in T^n$. J is the measurement matrix. $T(l)$ is used as the covariance matrix of measurement noise $B(l)$.

The above Formulas (5) and (8) are the mathematical model and observation model of the state matrix in the three-dimensional coordinate system of the joint points of the human body. Among them, the joint point models are the position and velocity components under the X, Y, and Z axes. $X_o(l)$, $E(l)$, $Z_o(l)$, $B(l)$, etc. are the results in six dimensions. In the actual solution, the X, Y, and Z axes are independent and similar to each other, which can be calculated under the X, Y, and Z axes, respectively, by using the dimensionality reduction method. Then, the estimated value of the current skeleton point in the x dimension is as follows:

$$x_o(l) = \begin{bmatrix} x_o(l) \\ \dot{x}_o(l) \end{bmatrix} = \begin{bmatrix} 1 & \Delta u \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_o(l-1) \\ \dot{x}_o(l-1) \end{bmatrix} + \begin{bmatrix} e_{x_o}(l) \\ e_{\dot{x}_o}(l) \end{bmatrix} \tag{10}$$

$$z_o(l) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_o(l) \\ \dot{x}_o(l) \end{bmatrix} + b_o(l) \tag{11}$$

With extensive data collection, $S = \begin{bmatrix} 1 & \Delta u \\ 0 & 1 \end{bmatrix}$, $J = [1 \quad 0]$, $W(l)$ and $T(l)$ can be derived. Through multiple data acquisition and calculation, $W(l)$ is a diagonal matrix with a value of 0.25 on the diagonal. $T(l)$ Is a diagonal matrix with values three on the diagonal. Then, the Kalman gain matrix L_l of the intermediate operation is derived from this formula. The estimated position of the skeletal joint point o in l frame time is obtained.

C. Trainer Gesture Recognition

In dance training, a trainer's posture assessment is an important factor in measuring the standard of his movements. The images produced by Kinect contain dance trainers and training scenes. Therefore, the target is extracted from the training object. In this paper, the inter-frame difference method is used to extract the moving objects of athletic dance trainers.

When using the inter-frame difference method to extract the moving target of the athletic dance trainer, it is necessary to compare the two frames of images and set the threshold. If the difference between two adjacent images is less than a threshold, the content of this image is a training scene, not a training object. If the difference between two adjacent images is greater than a threshold, then the content of this image is the trained object. This program can be expressed as:

$$Q(x, y) = |O_{l+1}(x, y) - O_l(x, y)| \tag{12}$$

$$F(x, y) = \begin{cases} 0, Q(x, y) < U \\ 255, Q(x, y) \geq U \end{cases} \tag{13}$$

In the formula, $Q(x, y)$ is the difference image of $O_{l+1}(x, y)$ and $O_l(x, y)$ for two adjacent two frames. $F(x, y)$ To $Q(x, y)$ binarized image. U is the set threshold.

After identifying the training object, feature extraction is performed. In this paper, the Canny algorithm is used to extract the features of the training objects. The Canny algorithm detects the boundaries of the objects being trained. Then, feature extraction is performed by connecting edges. When Canny's algorithm extracts features, it must smooth them:

$$G(x, y) = \frac{1}{2\pi\delta^2} exp\left(-\frac{x^2+y^2}{2\delta^2}\right) \tag{14}$$

Next, in the image (x, y), the gradient magnitude $Z(x, y)$ and the direction $f(x, y)$ on the pixel point are found :

$$Z(x, y) = \sqrt{a_x^2(x, y) + a_y^2(x, y)} \tag{15}$$

$$f(x, y) = arv\ tan\left(\frac{a_x(x,y)}{a_y(x,y)}\right) \tag{16}$$

Then, the non-maximum denoising of gradient magnitude $Z(x, y)$ and direction $f(x, y)$ are used to get candidate edges. An edge point extraction method based on the double-threshold method is proposed and stitched to realize the feature extraction of images.

In response to this situation, a gradient direction-based suppression algorithm is proposed, which is used to extract and connect candidate edge points.

By detecting human body points, the location of the human body can be realized through the point information in the sample library. This paper uses the Euclidean distance method to perform the matching operation between nodes. The specific steps are as follows:

$$J = \sqrt{(x_1, x_2)^2 + (y_1, y_2)^2} \tag{17}$$

In the formula, J is used to represent the corresponding pose, and the node coordinates and joint point coordinates detected in the sample library are (x_1, x_2) and (y_1, y_2).

Using the above steps, the body poses estimation of the sports dance trainer is completed.

III. Experiment of Dance-Assisted Training based on Kinect

In light of Kinect, this paper fosters a dance partner preparing framework in view of Kinect. This paper utilizes Kinect innovation to gather standard movement information for a PC. A data set of standard dance acts is made and utilized format. The mentor's development information and the layout pose are contrasted to understand the assessment of the learner's development pose, which brings the digitalization of dance preparation into another stage.

Ballet is the most logical and standard dance. It focuses on the accuracy of the developments and communicates the artists' feelings and contemplations with exclusive requirement dance steps. Expressive dance essential preparation is an expert, standard, and logical dance education. To this end, this article treats fundamental artful dance preparation as a trial dance-helped workout.

A. System Design

The centre substance of the framework incorporates information securing, handling, and examination. On this premise, the movement boundaries are investigated. A set of standard dance developments is built using the mentor's development data. The information handling is, for the most part, to fix the impeded joints and re-establish them. Information examination is to think about the development data between the mentor and the standard developments. The related preparation plan is acquired in order to work on the capacity of the mentor rapidly. The engineering of the framework is displayed in Fig. 6.

Hardware for this article includes Kinect for Microsoft Windows 2.0 and a PC with Windows 8. The Kinect is connected to the computer using a USB interface. Kinect is used to acquire 2D, deep images of the human body, which are also analyzed and processed.

In terms of software, it mainly includes capturing the action and analyzing action information. The motion information collection part mainly completes the collection of two-dimensional images of the human body, skeletal nodes, and depth. The movement information analysis compares the couch's motion information with the corresponding data in the confirmed dance motion database. The key point is to compare each link's positions and joint angles. Fig. 7 shows the operation interface design of the system.

B. Data Collection

To justify the dance steps of the students, it is necessary to have a standard dance step for comparison. This part uses the trainer's action extraction and the occlusion point information to restore and obtain the standard motion data.

This article maps each data label to an action name. It is stored uniformly as exercise information for training personnel to compare. There is a whole of 30 sets of primary instructing moves in the info archive. Every group is divided into four decomposing movements, with a total of 120 movements, which meets the basic requirements of the students.

Using Kinect technology, data collection of ballet hand positions is carried out. The trainer performs a series of dance steps before the Kinect. Kinect can capture the trainer's dance in real time and correct it through the occlusion point to obtain the trainer's physical condition. Finally, the training results are compared with the standard motion data.

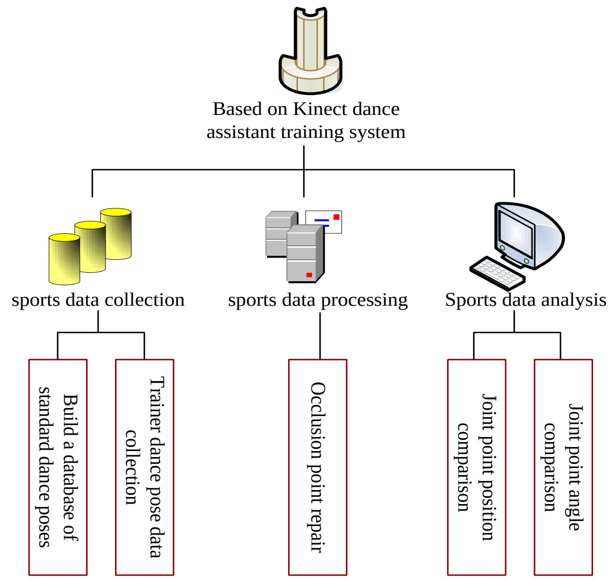


Fig. 6. Block diagram of the Kinect-based dance assistant training system.

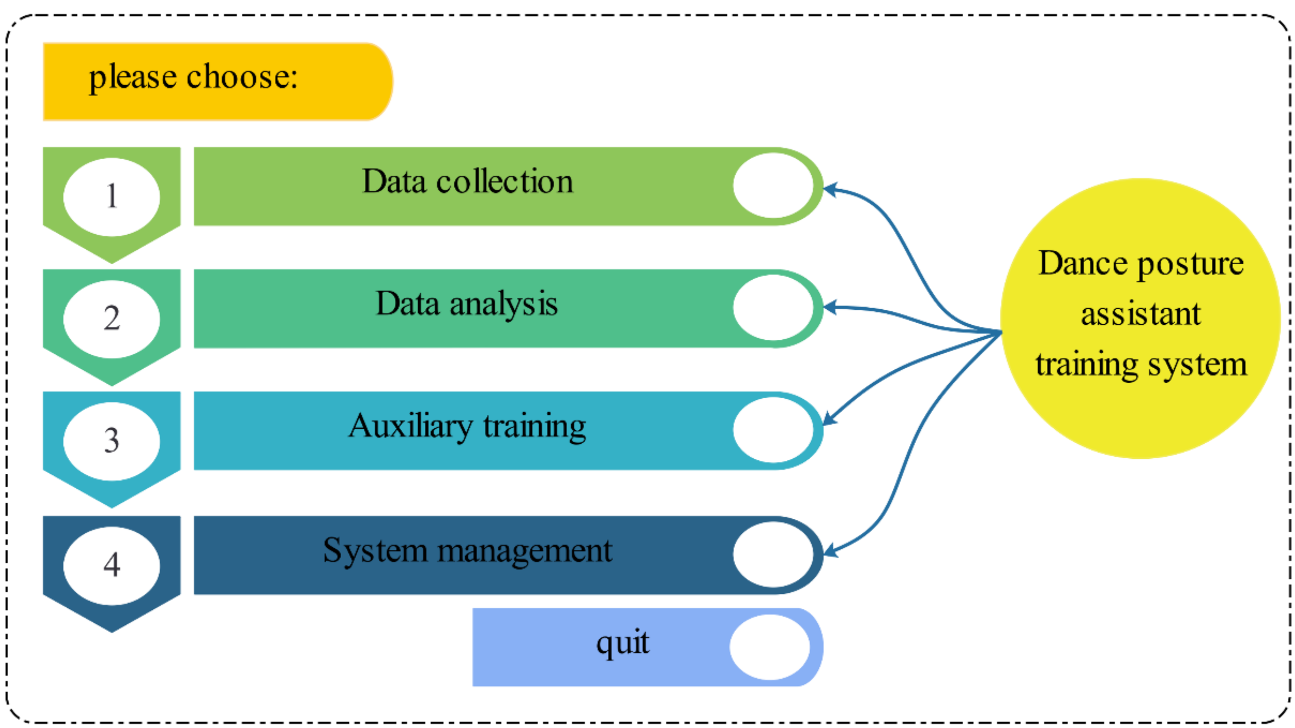


Fig. 7. Design of the software interface of the auxiliary training system.

C. Dance Auxiliary Training

This section of dance-assisted training mainly analyzes the trainer's key parts and the joints' angles. The auxiliary training system (see Tables I and II) collects each trainee's movement key point coordinates, which are compared with the standard movements of the choreographer (see Table III). By comparing the movement trajectories of various points, it is easy to find the dissimilarity between the couch's move and the approved move.

TABLE I. COUCH 1 ACTION JOINT COORDINATES

Part	Standard dance joint point coordinates	
	Abscissa	***Y-axis***
head	-79.2	217.1
Neck	-24.6	31.7
left shoulder	-149.9	-49.1
left elbow	-376.1	-51.8
left hand	-652.9	-60.2
Right hand	96.5	-7.4
Right elbow	81.4	222.7
Right shoulder	48.0	508.4
Right knee	-106.4	-864.6
Right foot	-67.0	-1158.7
left knee	-236.7	-936.7
left foot	-312.5	-1169.7

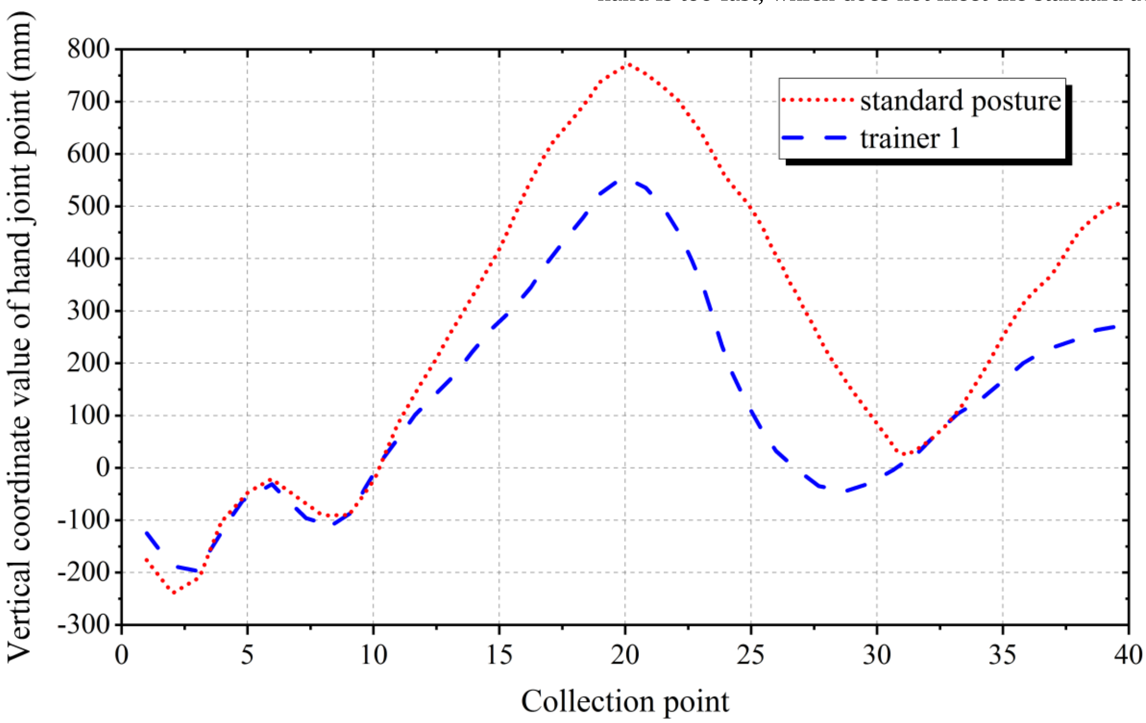
TABLE II. TRAINER 2 JOINTS ACTION COORDINATES

Part	Standard joint dance spot coordinates	
	Abscissa	***Y-axis***
head	-81.4	217.2
Neck	-26.6	31.0
left shoulder	-151.5	-51.3
left elbow	-376.1	-53.6
left hand	-652.8	-62.5
Right hand	96.8	-9.6
Right elbow	80.9	222.7
Right shoulder	47.4	508.1
Right knee	-108.0	-866.4
Right foot	-68.8	-1160.2
left knee	-238.5	-938.4
left foot	-314.6	-1172.0

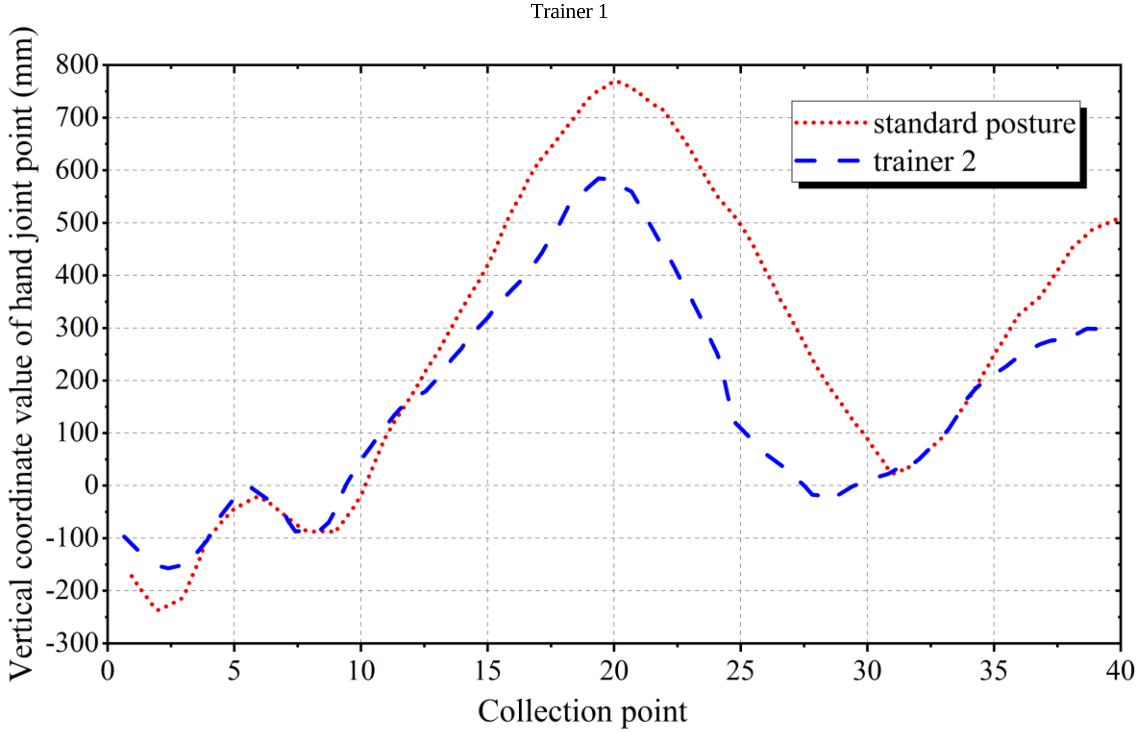
TABLE III. STANDARD ACTION JOINT COORDINATES

Part	The trainer's joint spot coordinates	
	Abscissa	***Y-axis***
head	-68.5	269.9
Neck	-68.5	74.9
left shoulder	-208.7	-2.4
left elbow	-407.7	-108.5
left hand	-634.8	-180.2
Right hand	309.9	373.1
Right elbow	237.0	153.2
Right shoulder	91.2	1.6
Right knee	-127.4	-828.1
Right foot	-161.0	-994.5
left knee	-23.6	-856.9
left foot	-12.4	-1000.3

It can be seen from Table I, Table II, and Table III that the height of the trainer's arm raises to the highest point does not meet the requirements. The difference between the data of Trainer 1 and Trainer 2 is not particularly large, which is within a certain range. It can be seen that the abscissa of the coordinates of the two trainers is slightly different from the abscissa of the standard dance joint point. The standard abscissa is -12.4. Trainer 1 is -312.5, and Trainer 2 is -314.6, which is over 300 errors. The comparison of Fig. 8 can be drawn from the above data. As can be seen from the graphs in Fig. 8(a) and 8(b), the trainer's wrist should be raised by approximately 200 mm. Moreover, the speed of closing the hand is too fast, which does not meet the standard action.



Trainer 1



Trainer 2

Fig. 8. Comparison of the movement trajectories of the two trainers and the standard movement trajectories.

On this basis, three hand movements in dance are analyzed. The included angle of each joint in this action is obtained. It can be seen that there is a certain difference between the movements of the trainer and the standard movements. Teachers of different ranks would be able to carry out aimed teaching based on their own abilities. Fig. 9 shows the disparity of the joint angles between the trainer and the standard movement. Fig. 9(a), (b), and (c) are the comparisons between Action 1, Action 2, and Action 3, respectively.

The following description is made with reference to operation 1 (operation at five positions) of Fig. 9(a). As can be seen from the figure, the angle between the trainer's right wrist-right elbow-right shoulder is rather big. The included angle of the waist-right knee-right ankle is also rather big. Therefore, the next time the trainer exercises, the five elbows and right arms should be straightened. At the same time, the trainer should lean the waist slightly forward and bend the right leg. The angle between the neck-right shoulder-right elbow is too big, and the angle between the left wrist-left elbow-left shoulder is relatively tiny. Therefore, in the next exercise, the trainer should try to reduce the curvature of the left elbow and left wrist as much as possible.

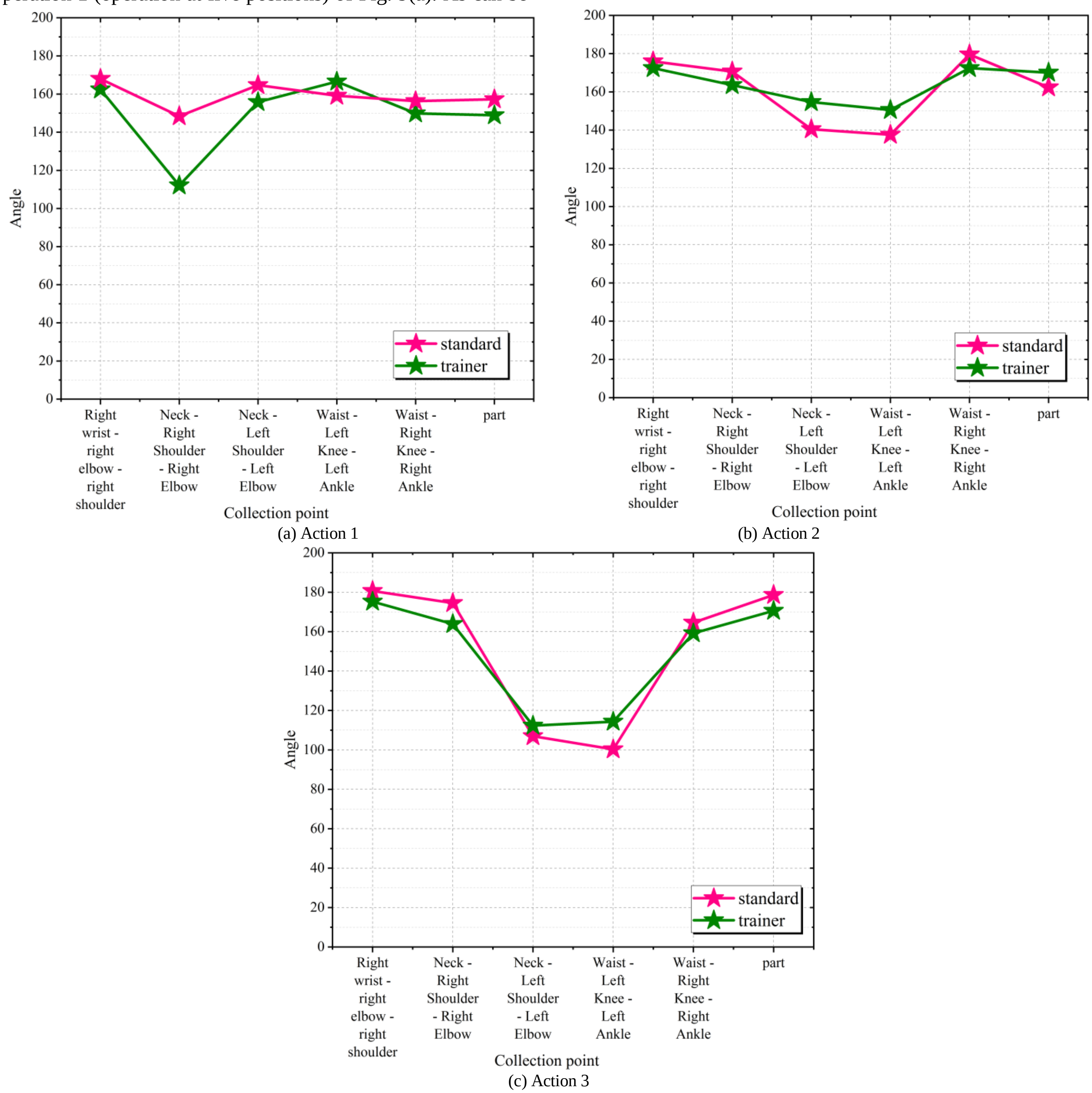
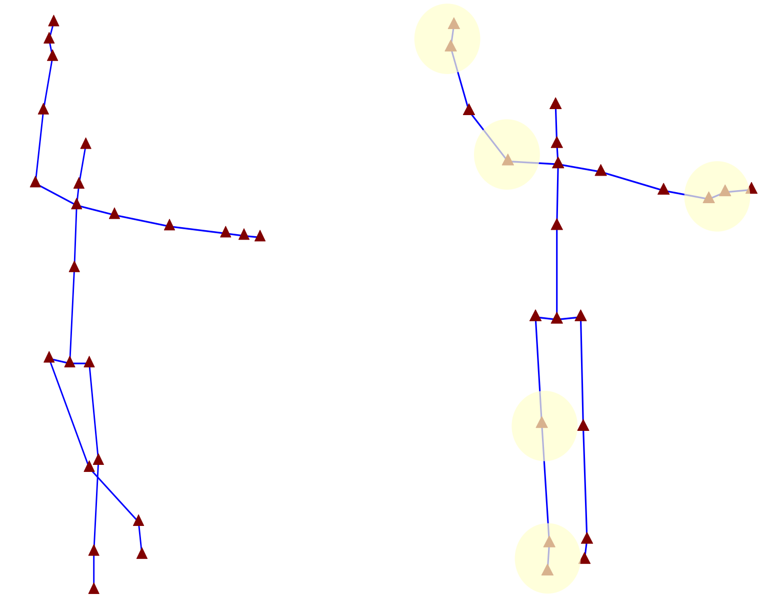


Fig. 9. Comparison of the joint angles of the trainer and the standard movement.



(a) Trainer's five-position hand movement (b) Standard five-position hand movement

Fig. 10. Comparison of the five-hand movements of the trainer and the standard five-hand movements.

In order to more intuitively compare the movements of the trainer and the dance instructor, the system shows the dissimilarity between the teacher's and the approved moves, as shown in Fig. 10. Fig. 10(a) is the operation of five hand exercise trainers. Fig. 10(b) is a standard five-hand exercise.

The equipment and programming portions of the framework are talked about exhaustively above as far as information assortment and standard dance moves are placed into the framework. A standard human dance present data set is built. The information handling part fixes the covered key parts and re-establishes the skeleton of the human body. Assistant preparation incorporates joint point situating and joint point-based helper preparation. The artist's dance act is examined by the directions of the endpoint and the point framed by the joints. The framework can acquire the related preparing program through correlation, which empowers understudies to make fitting changes as indicated by the prompts of the product in order to accomplish the impact of helper preparing.

IV. Conclusion

In this study, we leveraged the capabilities of the Kinect3D sensor and harnessed its skeletal tracking technology to capture comprehensive dance movements and training data. These data were meticulously compared to standard movements, providing users with a valuable tool to discern the nuances of their own performance. This innovative approach has far-reaching implications for dance instructors, professional dancers, and dance enthusiasts seeking to refine their skills through dedicated training and self-education.

The primary achievement of our work lies in its capacity to empower individuals to make precise adjustments to their dance postures, thereby ensuring the utmost accuracy in their dance practice. A significant advantage of this method is the liberation from the constraints of geographical location, time limitations, and spatial boundaries. Dancers can now engage in training sessions at their convenience, transcending traditional barriers. This expanded accessibility to training opportunities enhances the versatility and adaptability of our dance motion detection algorithm.

Furthermore, our system offers an interactive experience that not only facilitates correct posture adjustments but also cultivates a sense of rhythm among users through computer-assisted feedback. By embracing digitalization and informatization, we contribute to the ongoing modernization of dance training methods.

In conclusion, our work demonstrates the potential of computer vision and Kinect technology to revolutionize dance training. The ability to receive real-time feedback and guidance, combined with the freedom of practice, opens new avenues for dancers to hone their skills. However, it is essential to acknowledge that our approach does have limitations, including [mention limitations here], and we envision several exciting directions for future research. As we continue to explore these frontiers, we anticipate further advancements in the field of dance motion detection and its application in enhancing dance education and practice.

Acknowledgment

No funding was used to support this study.

Data Availability

Data is available in our manuscript.

Conflicts of Interest

We confirm there are no potential competing interests in our manuscript.

References

[1] E. Skoe, E. V Scarpati, and A. McVeety, “Auditory temporal processing in dancers,” *Percept Mot Skills*, vol. 128, no. 4, pp. 1337–1353, 2021.

[2] C. Thornquist, “The potential of dance: Reducing fashion consumption through movement therapy,” *J Clean Prod*, vol. 183, pp. 824–830, 2018.

[3] J. S. K. Reddy and C. Pereira, "From expressions to ecstasy: understanding the phenomenon of experiential interaction between the performer and audience in dance," *Dance, Movement & Spiritualities*, vol. 5, no. 1, pp. 89–99, 2018.

[4] G. Sun, W. Chen, H. Li, Q. Sun, M. Kyan, and P. Zhang, "A Virtual Reality Dance Self-learning Framework using Laban Movement Analysis.," *Journal of Engineering Science & Technology Review*, vol. 10, no. 5, 2017.

[5] T. Shiratori, A. Nakazawa, and K. Ikeuchi, "Detecting dance motion structure through music analysis," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, 2004, pp. 857–862. doi: 10.1109/AFGR.2004.1301641.

[6] X. Chang, Z. Ma, M. Lin, Y. Yang, and A. G. Hauptmann, "Feature Interaction Augmented Sparse Learning for Fast Kinect Motion Detection," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3911–3920, 2017, doi: 10.1109/TIP.2017.2708506.

[7] A. Barbu, Y. She, L. Ding, and G. Gramajo, "Feature selection with annealing for computer vision and big data learning," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 2, pp. 272–286, 2016.

[8] Y.-J. Cha, J. G. Chen, and O. Büyüköztürk, "Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters," *Eng Struct*, vol. 132, pp. 300–313, 2017.

[9] S. Khan, H. Rahmani, S. A. A. Shah, M. Bennamoun, G. Medioni, and S. Dickinson, *A guide to convolutional neural networks for computer vision*, vol. 8, no. 1. Springer, 2018.

[10] W. Fang, L. Ding, H. Luo, and P. E. D. Love, "Falls from heights: A computer vision-based approach for safety harness detection," *Autom Constr*, vol. 91, pp. 53–61, 2018.

[11] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, May 2017, doi: 10.1111/mice.12263.

[12] D. Roberts, T. Bretl, and M. Golparvar-Fard, "Detecting and Classifying Cranes Using Camera-Equipped UAVs for Monitoring Crane-Related Safety Hazards," in *Computing in Civil Engineering 2017*, pp. 442–449. doi: 10.1061/9780784480847.055.

[13] L. Ding, W. Fang, H. Luo, P. E. D. Love, B. Zhong, and X. Ouyang, "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory," *Autom Constr*, vol. 86, pp. 118–124, Feb. 2018, doi: 10.1016/j.autcon.2017.11.002.

[14] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[15] P. Pareek and A. Thakkar, "A survey on video-based human action recognition: recent updates, datasets, challenges, and applications," *Artif Intell Rev*, vol. 54, pp. 2259–2322, 2021.

[16] A. Dixit, R. Kumar Chidambaram, and Z. Allam, "Safety and risk analysis of autonomous vehicles using computer vision and neural networks," *Vehicles*, vol. 3, no. 3, pp. 595–617, 2021.

[17] A. Sööt and E. Viskus, "Contemporary approaches to dance pedagogy–The challenges of the 21st century," *Procedia-Social and Behavioral Sciences*, vol. 112, pp. 290–299, 2014.

[18] J. Jin and B. Snook, "Comprehensively strengthening and improving aesthetic education in a new era: An examination of the dance education major at the Beijing dance academy," *International Journal of Chinese Education*, vol. 11, no. 3, p. 2212585X221127451, 2022.

[19] L. Wang, D. Q. Huynh, and P. Koniusz, "A comparative review of recent kinect-based action recognition algorithms," *IEEE Transactions on Image Processing*, vol. 29, pp. 15–28, 2019.

[20] Y.-F. Xiao, H. Huang, J. Zheng, and L. Ran, "Obstacle detection for robot based on kinect and 2d lidar," *Journal of University of Electronic Science & Technology of China*, vol. 47, no. 3, pp. 337–342, 2018.

[21] H. Liu, C. Pan, Y. Shen, and B. Gao, "Plant Point Cloud Information Fusion Method Based on SICK and Kinect Sensors," *Trans. Chin. Soc. Agric. Mach*, vol. 49, no. 10, pp. 284–291, 2018.

[22] Y. Sun *et al.*, "Retracted: Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images," *IET Image Process*, vol. 14, no. 15, pp. 3662–3668, 2020.

[23] Y. Z. Xin and Z. F. Xing, "Human action recognition method based on Kinect," *Computer Engineering & Design*, vol. 37, no. 4, pp. 1056–1061, 2016.

[24] X. Wang, J. Liu, and Q. Zhou, "Real-time multi-target localization from unmanned aerial vehicles," *Sensors*, vol. 17, no. 1, p. 33, 2016.