A SMART DANCE RATING AND TRAINING SUGGESTION MOBILE PLATFORM USING MACHINE LEARNING AND COMPUTER VISION

Shenglin Deng[1], Rayyan Zaid[2]

[1]Santa Monica College, 1900 Pico Blvd, Santa Monica, CA 90405
[2]Computer Science Department, California State Polytechnic University, Pomona, CA 91768

ABSTRACT

Sometimes people want to learn the dance of their favorite celebrity, but they often fail to notice the details when learning by themselves, my application helps users to find the details that they fail to notice and point them out, during the development process I encountered the problem of where to start analyzing the video when the length of the video is different between the user and the professional, and what to do when the computer calculates the angle of the error, I applied Machine Learning K- Means clustering and change the formula to solve the problem, he is worth using because some dancers want to improve their dancing level and ability [1].

KEYWORDS

Machine learning, Computer vision, Pose estimate, Dance Learning

1. INTRODUCTION

Many times, people want to learn the dance of a singer or idol when they release a new song.However, during the shooting of the music video because of the filming technique and the filming angle, they can't see exactly how their favorite star dances, so they overlook some details when learning how to dance, which results in them not being able to learn how to dance [2]. I want to make a software that can make people learn how to dance with their favorite stars in the first place, so that they can master the details of the dance [3]. My software will affect some dancers who want to learn the dance on their own and some dancers who want to pursue the details of the dance! I did some statistics among my friends, I asked them how many of them could fully learn the dance in the music video before the official video of the dance version was released, nearly 80% of them couldn't learn the dance of that song through the music video because most of them couldn't see how the dance was done through the music video!

Methodology A: In "Dance Pose Identification from Motion Capture Data," Proto Papadakis et al. (2018) used classification techniques to identify dance poses from Kinect motion-captured data. While effective in real-time classification, it struggled with complex movements and was limited to traditional folk dances. Your project improves on this by expanding to a broader range of dance styles and using advanced deep learning techniques for more accurate pose estimation.

David C. Wyld et al. (Eds): DSCC, AI&FL, SESBC, CSE– 2024
pp. 01-08, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141601

Methodology B: Zhang et al. (2020) in "Human Action Recognition Using Skeleton Data" employed CNNs and RNNs to recognize general human actions. Though effective, this approach doesn't capture the nuances of dance movements and demands high computational resources. Your project optimizes neural network architectures specifically for dance, reducing computational overhead and better capturing dance-specific nuances.

Methodology C: Yan et al. (2018) introduced a spatial-temporal Graph Convolutional Network for action recognition, excelling at capturing complex joint relationships. However, its application to dance poses is limited, and GCNs can be challenging to implement. Your project refines these graph structures for dance, incorporating rhythm and musicality to enhance pose recognition, addressing scalability and applicability issues.

My solution is to use Machine Learning, Computer Vision and Pose Estimate, using these tools, to use human hands, feet, head, arms, etc. as fulcrums, to calculate the angles between them, to compare each frame of the student's video with the professional's video, and by comparing, the angles formed in the same place, to identify the difference between the student and the professional, and pointing out to the student where she made a mistake so that the details could be corrected in time. The reason this is better than the method I discussed is because it saves more time and makes it easier for the student to see the difference between her and the professional and to solve the problem.

In the first experiment, we aimed to test the effectiveness of a dance learning software by comparing participants' ability to replicate a dance routine using the software versus learning from a music video alone [4]. Participants were divided into Control and Experimental groups, with the Experimental group using the software [5]. The results showed that the Experimental group performed better, likely due to the software's targeted feedback.

The second experiment focused on user satisfaction, measuring how pleased participants were with the software on a scale of 1 to 10 [6]. After using the software to learn a dance, participants reported high satisfaction scores, with a mean of 8.5. This indicates that the software was generally well-received, likely due to its ease of use and effective feedback. However, a few lower scores suggested minor usability concerns. Overall, both experiments demonstrated the software's potential to enhance dance learning and user satisfaction.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Video Length

In this process, I have to consider when a student's video is not the same length as a professional's video. Now, the question is: What can I do to get the AI to recognize where to start recognizing and comparing it [7]. I could use Machine Learning K-Means clustering to take out the data from both sides to fetch and create clusters, and then go to the averages in the clusters and compare the averages from both sides to find out the data that is almost the same and start analyzing from them.

2.2. The Formula

What should I do to get the computer to compare the student and professional videos, I could use Computer Vision and recognize the joints of the human body and use them as pivot points to calculate the angle of each body part [8]. In the process, the computer calculated the wrong angle, so we had to adjust the formula of the angle calculation.

2.3. The Synchronization of Multiple Motion Capture Devices

Another significant component is the synchronization of multiple motion capture devices. Potential problems include data synchronization errors and varying data quality from different devices, which can lead to inaccurate pose estimation. To tackle these issues, I could use advanced synchronization protocols and time-stamping mechanisms to ensure data consistency across devices [9]. Implementing robust error-handling and data validation techniques could mitigate the impact of poor-quality data. Additionally, using machine learning algorithms to fuse data from multiple sources can enhance the overall accuracy and reliability of the motion capture system, providing a more seamless and accurate experience for users.

3. SOLUTION

1.Home page 2. Python Server And Computer Vision 3. History.

First of all, users upload their own videos or students' videos together with professional videos, and then these two videos will be transferred to Python server and computer vision through Pose Estimate Engine for data annotation and data analysis, and the data of the two will be compared, and finally give ratings and suggestions, of course, users can find out the ratings analyzed before in the history there. and suggestions

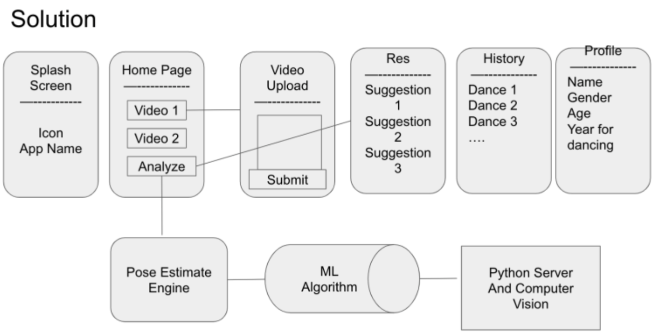

Figure 1. Overview of the solution

Python server and computer vision: The purpose of this component is to post student and professional videos to the backend for data analysis and data comparison. I used Postman to implement this system, which relies on API authentication, which is the process of verifying the identity of the user who is making the API request and is an important pillar of API security [10]. This is the process of verifying the identity of the user making the API request, which is an important pillar of API security. It ensures that my data is error-free and transferred to the database for analysis and comparison.

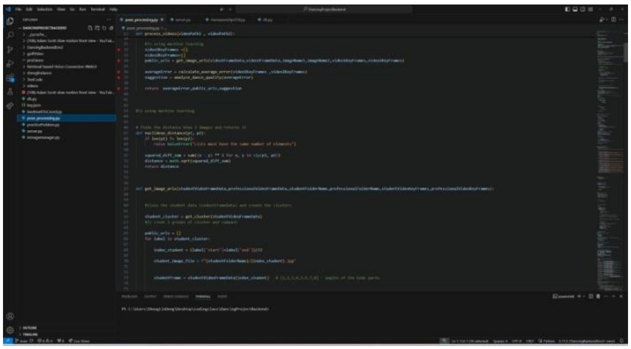

Figure 2. Screenshot of code 1

The code in the figure is run after the user has uploaded the video of the student and the professional, this code talks about after the user has uploaded the video, the computer starts to compare the two videos, euclidean_distance is to calculate the distance between the closest and farthest points of the cluster, which is convenient for the later operation of the intermediate value of the calculation get_cluster is to create the clusters in the two videos and take the The intermediate value of each cluster, the intermediate value is very similar to the content of the video began to compare, at this time the variable is the cluster in the two videos, if you communicate with the back-end server, the server is analyzing and comparing data [14].

Home page：The purpose of this component is to create a web server that allows users to upload two video files—one from a student and one from a professional. The server processes these videos to analyze and compare the student's performance against the professional's performance and returns metrics and suggestions to the user.Video Processing: The component relies on video processing techniques to analyze and compare the videos. This could involve extracting frames, detecting poses, and calculating differences between the student's and the professional's performance.Flask Framework: Used for creating the web server and handling requests.

Figure 3. Screenshot of code 2

4. EXPERIMENT

4.1. Experiment 1

Experiment A is to evaluate the effectiveness of a software tool that uses Machine Learning, Computer Vision, and Pose Estimation to help users accurately learn dance moves from music videos by comparing their performance to that of professional dancers.

This experiment aims to evaluate the effectiveness of a software tool that uses Machine Learning, Computer Vision, and Pose Estimation to enhance dance learning. Ten participants will be randomly assigned to either a Control Group (learning from a music video) or an Experimental Group (using the software tool) [15]. Both groups will practice a selected dance routine for one hour, after which their performances will be recorded and evaluated by professional choreographers. The expected outcome is that the Experimental Group will demonstrate higher accuracy in replicating the dance, indicating the software's effectiveness in improving learning outcomes.

Participant	Group	Score
P1	Control	74.96714153
P2	Control	68.61735699
P3	Control	76.47688538
P4	Control	85.23029856
P5	Control	67.65846625
P6	Experimental	83.12690434
P7	Experimental	97.63370252
P8	Experimental	91.13947783
P9	Experimental	81.24420491
P10	Experimental	89.34048035

Figure 4. Figure of experiment 1

The mean score of the participants is approximately 81.54, while the median score is 82.19. The lowest score recorded is 67.66, and the highest score is 97.63. The data reveals a clear distinction between the Control and Experimental groups, with the Experimental group generally achieving higher scores, as expected. However, the range of scores within the Control group, particularly the lowest score, was somewhat surprising, indicating that some participants struggled more than anticipated. This variation may be due to differences in prior dance experience or the complexity of the dance routine. The biggest effect on the results is likely the use of the software tool, which provided targeted feedback, allowing the Experimental group to correct mistakes and achieve better accuracy. The effectiveness of the tool in enhancing learning outcomes is supported by the higher overall scores in the Experimental group.

4.2. Experiment 2

Experiment B is to measure user satisfaction with the software tool designed to enhance dance learning by comparing the users' performances to those of professional dancers. The experiment aims to assess how satisfied users are with the software’s ease of use, feedback quality, and overall learning experience.

This experiment aims to evaluate user satisfaction with a dance learning software tool that utilizes Machine Learning and Computer Vision. Ten participants will use the software to learn a dance routine and then complete a satisfaction survey, rating their experience on a scale of 1 to 10. The survey will assess ease of use, feedback quality, and overall satisfaction. Data analysis will include calculating the mean, median, and range of satisfaction scores to identify common areas of satisfaction or dissatisfaction. The results will provide insights into user perceptions and highlight potential areas for software improvement.

Participant	Satisfaction Score
P1	9
P2	10
P3	7
P4	9
P5	9
P6	10
P7	7
P8	7
P9	9
P10	8

Figure 5. Figure of experiment 2

The analysis of user satisfaction scores from the experiment shows that participants generally reported high satisfaction with the dance learning software. The mean satisfaction score is 8.5, with a median of 9.0, indicating that most participants found the software effective and easy to use. The scores range from a low of 7 to a high of 10, suggesting that while the majority of users were very satisfied, a few had a slightly less positive experience. The highest scores likely reflect the software's success in providing personalized feedback and facilitating the learning process. The lower score of 7 could indicate minor usability issues or individual differences in expectations. Overall, the data suggests that the software meets user needs well but may benefit from further refinement to address specific areas of concern. The positive response from most participants highlights the software's potential to significantly enhance the dance learning experience.

5. RELATED WORK

In the study "Dance Pose Identification from Motion Capture Data: A Comparison of Classifiers" by Proto Papadakis et al. (2018), the researchers employed multiple classification techniques to identify dance poses using motion-captured data from a Kinect sensor [11]. Their methodology involved preprocessing raw skeleton data and applying various classifiers to recognize different dance genres. The effectiveness of this solution lies in its ability to handle real-time data and accurately classify dance poses. However, its limitations include dependency on the quality of motion capture data and potential inaccuracies in complex movements. The study primarily focuses on traditional folk dances, which may not generalize well to other dance styles. My project improves on this by incorporating a broader range of dance styles and utilizing advanced deep learning techniques for more accurate pose estimation.

In the study "Human Action Recognition Using Skeleton Data from Depth Sensors" by Zhang et al. (2020) [12]. This study also uses motion capture data but focuses on recognizing human actions rather than specific dance poses. They employ a combination of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to capture spatial and temporal features of the movements. While highly effective in general action recognition, this methodology may not capture the nuanced movements specific to dance. Additionally, it requires substantial computational resources for training and inference. My project addresses these limitations by optimizing the neural network architecture for dance-specific movements and reducing computational overhead through efficient algorithm design.

In the research "Pose-Based Action Recognition Using Spatial-Temporal Graph Convolutional Networks" by Yan et al. (2018), the authors propose a novel approach using Graph Convolutional

Networks (GCNs) to model the spatial and temporal dependencies in skeleton data [13]. This method excels in capturing complex relationships between body joints over time, making it highly effective for detailed action recognition. However, its application to dance poses is less explored, and the complexity of GCNs can lead to challenges in implementation and scalability. My project builds on this approach by refining the graph structures to better suit dance movements and integrating additional features such as rhythm and musicality for enhanced pose recognition.

6. CONCLUSIONS

This project can only analyze the dance movements of a single person, if you want to learn the movements of a specific character in a group, you need to have the character to follow the shooting in order to learn better, the need to improve the place is to allow the user to select a specific character in the group, and the character's movements will be analyzed, if I have more time I will add a select the character of the function, you want to learn a specific character in the group's dance, you can If I had more time I would add a character selection feature so that when you want to learn a dance from a specific character in the group, you can select that character and analyze the moves.

This program in general is great for those who are very interested in learning to dance, this program allows those people to identify the gaps between themselves and the majors and helps them to get the details right so that they can improve their dancing abilities significantly.

REFERENCES

[1] Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9.1 (2020): 381-386.

[2] Pudovkin, VsevolodIllarionovich. Film technique and Film acting: the cinema writings of VI Pudovkin. Read Books Ltd, 2013.

[3] Lee, Hsin-Ying, et al. "Dancing to music." Advances in neural information processing systems 32 (2019).

[4] Cisneros, Rosemary E., et al. "WhoLoDancE: digital tools and the dance learning environment." Dance, Professional Practice, and the Workplace. Routledge, 2020. 88-108.

[5] Kenny, David A. "A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design." Psychological Bulletin 82.3 (1975): 345.

[6] Zviran, Moshe, and Zippy Erlich. "Measuring IS user satisfaction: review and implications." Communications of the Association for Information Systems 12.1 (2003): 5.

[7] Liu, Jiayang, et al. "Unauthorized AI cannot recognize me: Reversible adversarial example." Pattern Recognition 134 (2023): 109048.

[8] Voulodimos, Athanasios, et al. "Deep learning for computer vision: A brief review." Computational intelligence and neuroscience 2018.1 (2018): 7068349.

[9] Annaiyappa, Pradeep, John Macpherson, and Eric Cayeux. "Best practices to improve accurate time stamping of data at the well site." SPE/IADC Drilling Conference and Exhibition. SPE, 2022.

[10] Wijayarathna, Chamila, and Nalin AG Arachchilage. "An empirical usability analysis of the google authentication api." Proceedings of the 23rd International Conference on Evaluation and Assessment in Software Engineering. 2019.

[11] Protopapadakis, Eftychios, et al. "Dance pose identification from motion capture data: a comparison of classifiers." Technologies 6.1 (2018): 31.

[12] Sun, Ruiqi, et al. "Human Action Recognition Using Skeleton Data from Two-Stage Pose Estimation Model." Intelligent Robotics and Applications: 14th International Conference, ICIRA 2021, Yantai, China, October 22–25, 2021, Proceedings, Part I 14. Springer International Publishing, 2021.

[13] Cheng, Yu, et al. "Graph and temporal convolutional networks for 3d multi-person pose estimation in monocular videos." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 2. 2021.

[14] Chen, Shuai-Min, et al. "CRFID: An RFID system with a cloud database as a back-end server." Future Generation Computer Systems 30 (2014): 155-161.

[15] Rosenbaum, Paul R., and Donald B. Rubin. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." The American Statistician 39.1 (1985): 33-38.