Dance performance evaluation using monocular video capturing device utilizing Dynamic Time Warping and Mediapipe BlazePose GHUM

Adithyan S Raj[1], Adithya Suresh[2], Kakumanu Aman Chowdary[3], Manda Rajesh Reddy[4], Anjali T[5], *Department of Computer Science & Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham,* Amritapuri, India, [1]amenu4cse17510@am.students.amrita.edu, [2]adithyas@am.students.amrita.edu, [3]kachowdary@am.students.amrita.edu, [4]mrajeshreddy@am.students.amrita.edu, [5]anjalit@am.amrita.edu

***Abstract*—The health benefits of dancing are tremendous; it boosts immunity and cognitive capabilities in children. It also helps prevent osteoporosis in older women. Unfortunately, one of the most common reasons for people to shy away from dance is the fear of embarrassment that comes with doing it wrong. Our goal with this project is to provide a platform to evaluate dance moves and help overcome this fear. In our application, we record the user's dance moves from a monocular video capturing device and calculate accuracy. We utilize Mediapipe BlazePose GHUM and Dynamic Time Warping respectively for pose estimation and accuracy calculation.**

***Keywords*—Computer Vision, Time Series Analysis, Human Pose Estimation, Dance Evaluation, Feature Extraction**

I. Introduction

Dance is a form of self-expression that is unrivaled in terms of the control, diversity, and versatility it employs. When we hear body language we see people in suits working in big corporations, yet the level of control they exhibit over their movements can mesmerize crowds in the millions. It not only helps improve physical wellness but also has a huge role to play in the mental wellbeing of its practitioners. Furthermore, it improves balanced and functional muscles which becomes more important with age.[3][6]

The health benefits of dancing are tremendous; it boosts immunity and cognitive capabilities in children. It also helps prevent osteoporosis in older women. Unfortunately, one of the most common reasons for people to shy away from dance is the fear of embarrassment that comes with doing it wrong. Our goal with this project is to provide a platform to evaluate dance moves and help overcome this fear.

Some of the applications we focused on for this project are given below:

- Ideal for people who want to learn to dance from the comfort of their homes.
- By evaluating their dance moves and scoring them in numeric terms they could set SMART goals (i.e., Specific, Measurable, Achievable, Relevant, Time-bound).
- With a further update to our application we could allow the user to share their dance with their friends.
- It can also help dancers to analyze and learn their favorite dance videos. The feedback mechanism (score and accuracy), allows the user to compete with themselves.

The challenges we faced during this project is shown below:

- Due to the COVID-19 pandemic, the team members were limited to communicating through online communication platforms which were not reliable because of frequent connectivity issues.
- The most difficult part of our project was to connect the various aspects of the project into a continuous flow and make it usable for the end user.
- We were new to some of the technologies that were required to build our project so we had to constantly learn throughout the project.
- We struggled to figure out how to approach the problem of scoring files with different time frames (we later found out dynamic time warping could solve this problem).

II. Related Works

We have drawn inspiration for Dance Evaluation from some very interesting papers [4][24][25]. We saw that Laban Movement analysis [4][25] and Dynamic Time Warping[24][25] could be used for this purpose. Laban Movement analysis uses a feature space that aims to capture four components (Body, Effort, Shape, Space), and can be subsequently used for motion comparison and evaluation. We can also use Dynamic Time Warping for this purpose of comparing patterns in varying time frames. It has applications in information retrieval [1], pattern matching trading systems [13] and it is better than Euclidean Distance Clustering method when it comes to time series data that is of varying lengths [9].

We found a lack of resources for a user to evaluate their dance moves using their regular laptops or webcams. We knew that the necessary algorithms for building such an application were already made and were waiting to be implemented. We used PyQT for our application since it would mean that the entire code is in python (since all the pose estimation and scoring algorithms were in python)[12][14][15]. Also, it was the ideal choice to work with OpenCV [16]. I would say it is more of an implementation gap than a research gap.

Dance as an artform traces its path back to the beginning of civilization and has always had a prominent role in the customs and beliefs of a community [18][19]. But sadly there is a lot of stigma around dancing even now and beginners especially feel embarrassed to dance in front of people, which is one of the best ways to get feedback. What we wanted to do was to ease the initial process of dancing with help of a scoring system till the user reaches a stage where they no longer feel this fear. Another reason is the enormous health benefits that dancing could bring in [8]. It is a form of functional training which improves balance and muscle strength. A lack of muscle strength and balance can lead to serious falls in seniors [3]. It can even help people with autism reduce some of the detrimental effects [7].

III. Proposed Method

First the user stands in front of the camera while the motion capture starts. The pose estimation module starts capturing the pose data from each frame while recording and saving it as a file. On further processing, the pose data is converted to csv files which can be used for the scoring function. Now the user has to select the dance video he is trying to replicate, as a video file. This file will also be processed further for the scoring function. When both videos are processed, the user can display the score of his dance session. The score is calculated using the Dynamic Time Warping algorithm which is shown below:

```
Algorithm 1: Dynamic Time Warping Algorithm

int DTWDistance(s: array [1..n], t: array [1..m], w: int) {
    DTW := array [0..n, 0..m]

    w := max(w, abs(n-m)) // adapt window size (*)

    for i := 0 to n
        for j:= 0 to m
            DTW[i, j] := infinity
    DTW[0, 0] := 0
    for i := 1 to n
        for j := max(1, i-w) to min(m, i+w)
            DTW[i, j] := 0

    for i := 1 to n
        for j := max(1, i-w) to min(m, i+w)
            cost := d(s[i], t[j])
            DTW[i, j] := cost +
                                minimum(DTW[i-1, j ],    // insertion
                                 DTW[i  , j-1],    // deletion
                                 DTW[i-1, j-1])    // match
    return DTW[n, m]
}
```

In time series analysis, dynamic time warping (DTW) is one of the algorithms for measuring similarity between two temporal sequences, which may vary in speed. DTW has been applied to temporal sequences of video, audio, and graphics data — indeed, any data that can be turned into a linear sequence can be analyzed with DTW.

The idea to compare arrays with different lengths is to build one-to-many and many-to-one matches so that the total distance can be minimized between the two.

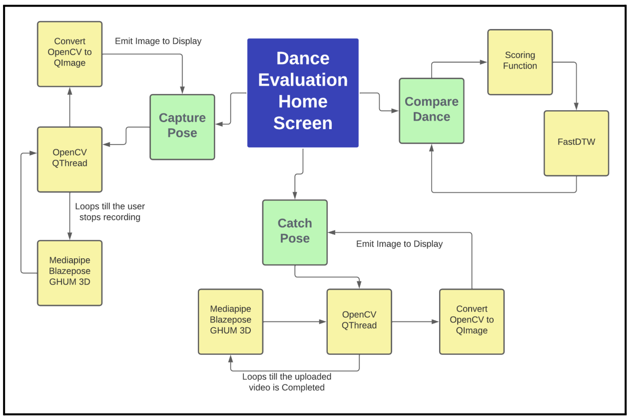

Fig 1: Block Diagram

The reason we chose Dynamic Time Warping algorithm against Euclidean distance is the difference in time between the two video files. By using Dynamic Time Warping we can compress the bigger file and compare the pattern to score the user.

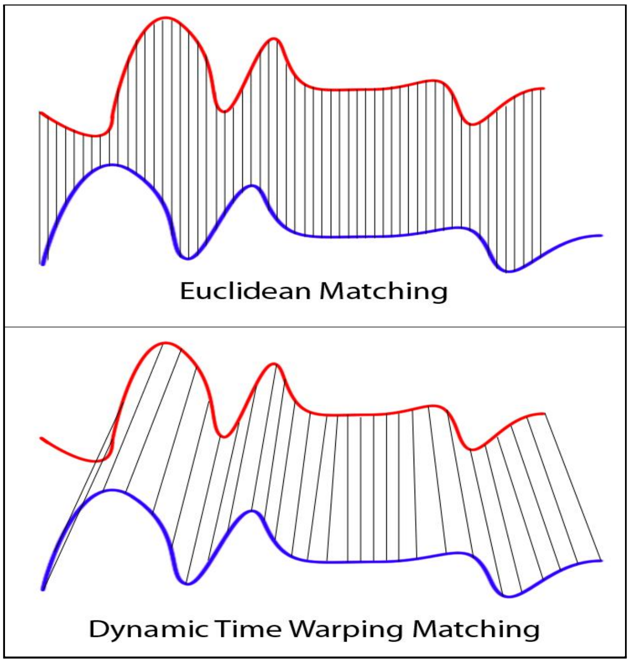

Fig 2: Euclidean Matching and Dynamic Time Warping Matching Comparison

We used Blazepose GHUM from Mediapipe library for our human pose estimation model. From which we selected 16 key points based on relevance to our model.

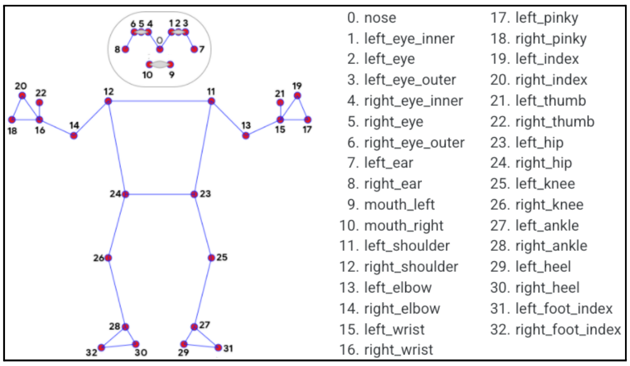

Fig 3: Pose Landmark Model (BlazePose GHUM 3D)

The various modules for our project is shown below:

- Capture pose
    - OpenCV Qthread
    - Mediapipe Blazepose GHUM 3D
    - Convert OpenCV to QImage
- Catch pose
    - OpenCV Qthread
    - Mediapipe Blazepose GHUM 3D
    - Convert OpenCV to QImage
- Compare dance
    - Scoring Function
    - FastDTW

Now let us discuss how these modules play a part in our project.

- **Catch pose** - Converts the user selected video to CSV file for further analysis
    - OpenCV QThread: This is the thread that starts the OpenCV VideoCapture and keeps calling Mediapipe Blazepose GHUM 3D till the entire video is finished.
    - Mediapipe Blazepose GHUM 3D: This algorithm is used to derive the body key points from the video file.

- Convert OpenCV to QImage: Here the image gets converted into PyQT5 readable format and gets emitted to the main thread.

- **Capture pose** - Records and converts the user's dance video for further analysis
  - OpenCV QThread: This is the thread that starts the OpenCV VideoCapture and keeps calling Mediapipe Blazepose GHUM 3D till the user stops recording.
  - Mediapipe Blazepose GHUM 3D: This algorithm is used to derive the body key points from the video file.
  - Convert OpenCV to QImage: Here the image gets converted into PyQT5 readable format and gets emitted to the main thread.
- **Compare Dance** - Compares and scores the performance of the user based on accuracy
  - Scoring Function: Here the scoring function calls the FastDTW function and gets the distance and path as return values. Using these values we calculate the accuracy and score and it is shown to the user.
  - FastDTW: We use this module for Dynamic Time Warping

IV. Results

FastDTW is an approximation of DTW which does a good job when it comes to larger time series data. So when the length of the time series increases the better it gets. We used FastDTW instead of DTW because dance videos can have a longer time series. We set the radius as the Euclidean distance of the two to datasets at that time.

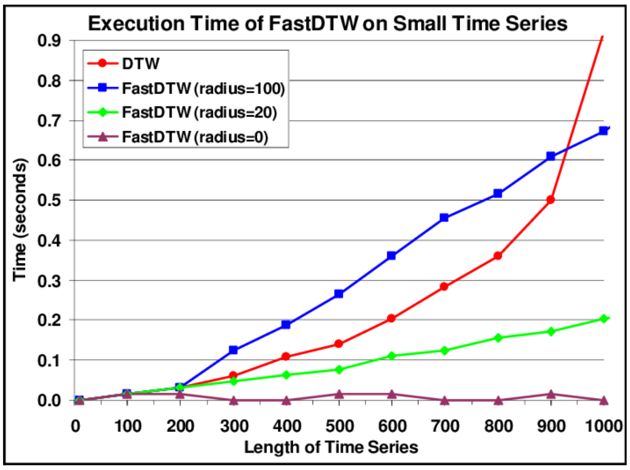

Fig 4: FastDTW vs DTW

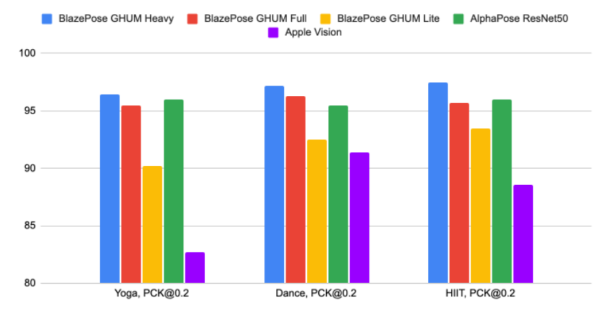

Fig 5: Quality Evaluation in PCK@0.2 (Percentage of Correct key-points)

When compared to AlphaPose ResNet50, Apple Vision and BlazePose GHUM Lite it has a better accuracy of about 96.3% when there are Common Objects in the frame. It even works at 67.4% accuracy when there are other objects on the screen. We could increase the accuracy to about 73% with GHUM Heavy but it would require a higher processing power and it had to be a balance between the two.

Method	Yoga mAP	Yoga PCK@0.2	Dance mAP	Dance PCK@0.2
BlazePose GHUM Heavy	68.1	**96.4**	73.0	**97.2**
BlazePose GHUM Full	62.6	**95.5**	67.4	**96.3**
BlazePose GHUM Lite	45.0	**90.2**	53.6	**92.5**
AlphaPose ResNet50	63.4	**96.0**	57.8	**95.5**
Apple Vision	32.8	**82.7**	36.4	**91.4**

Fig 6: Human Pose Algorithms Compared

The graph shown below compares the uploaded Video file and the recorded dance moves and plots a path based on the similarity of the two time series data. This helps us calculate the final accuracy with the following formula:

accuracy = round(((1 - (distance/median_total)) * 100), 2)

Where, median_total = median of path one + Median of path two, and distance is absolute value of Median of path one - Median of path two

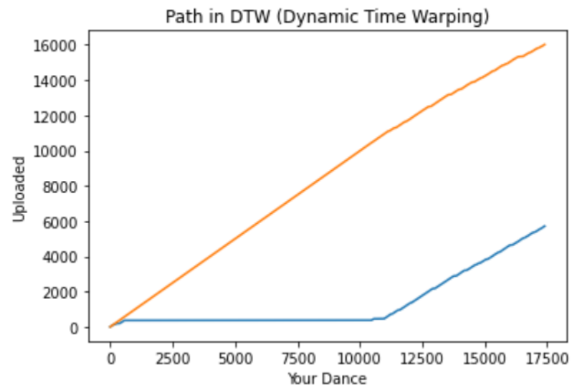

Fig 7: Comparison between Uploaded Video file and Recorded Dance Moves using DTW

V. Conclusion and Future Work

To sum it all up we discussed the various reasons why we took dance evaluation as our topic. We conveyed the various health and mental benefits of dancing and why it is more relevant now than ever before. We proposed a method to solve the problem by utilizing Mediapipe Blazepose GHUM 3D for Human pose estimation which returns us the necessary body key points for further analysis. We also focused on the scoring function which compares two datasets of different time lengths and compares them to return us the euclidean distance. We use Dynamic Time Warping for this purpose. Ultimately, the end goal for us is to make our product more convenient for the end user. Here are some future prospects that could make our product even better.

Future Prospects:

- Build on the application by bringing in statistical data on historical scores, achievements and streaks to motivate the user to keep coming back.
- Improve the speed of our algorithm using parallel programming and also utilize a separate thread for the Scoring function.
- Utilize better hardware and write more hardware centric code to speed up the system.

- Create a cross-platform application that can reach a bigger audience.
- Monetize to deploy the application and use premium memberships (or advertisements) for further development.

References

[1] Muller, Meinard. "Dynamic Time Warping in Information Retrieval for Music and Motion." (2007): 69-84.

[2] Bazarevsky, Valentin, et al. "Blazepose: On-device real-time body pose tracking." arXiv preprint arXiv:2006.10204 (2020).

[3] Granacher, Urs, et al. "The importance of trunk muscle strength for balance, functional performance, and fall prevention in seniors: a systematic review." Sports medicine 43.7 (2013): 627-641.

[4] Aristidou, Andreas, et al. "Folk dance evaluation using laban movement analysis." Journal on Computing and Cultural Heritage (JOCCH) 8.4 (2015): 1-19.

[5] Krasnow, Donna, and Steven J. Chatfield. "Development of the “performance competence evaluation measure”: assessing qualitative aspects of dance performance." Journal of Dance Medicine & Science 13.4 (2009): 101-107.

[6] Lin, Tzu-Wei, et al. "Different types of exercise induce differential effects on neuronal adaptations and memory performance." Neurobiology of learning and memory 97.1 (2012): 140-147.

[7] DeJesus, Beatriz Menezes, et al. "Dance promotes positive benefits for negative symptoms in autism spectrum disorder (ASD): A systematic review." Complementary therapies in medicine 49 (2020): 102299.

[8] LaMotte, Megan. "The integrated approach versus the traditional approach: Analyzing the benefits of a dance and transportation integrated curriculum." Journal of Dance Education 18.1 (2018): 23-32.

[9] Wang, Weizeng, et al. "Time series clustering based on dynamic time warping." 2018 IEEE 9th international conference on software engineering and service science (ICSESS). IEEE, 2018.

[10] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." arXiv preprint arXiv:2006.10214 (2020).

[11] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).

Meier, Burkhard. Python GUI Programming Cookbook: Develop functional and responsive user interfaces with tkinter and PyQt5. Packt Publishing Ltd, 2019.

[12] Kim, Sang Hyuk, et al. "Pattern matching trading system based on the dynamic time warping algorithm." Sustainability 10.12 (2018): 4641.

[13] Willman, Joshua. "Overview of pyqt5." Modern PyQt. Apress, Berkeley, CA, 2021. 1-42.

[14] Willman, Joshua M. "Managing Threads." Beginning PyQt. Apress, Berkeley, CA, 2022. 437-450.

[15] Willman, Joshua. "GUIs for Computer Vision." Modern PyQt. Apress, Berkeley, CA, 2021. 163-208.

[16] Shajina, T., and P. Bagavathi Sivakumar. "Human gait recognition and classification using time series shapelets." 2012 international conference on advances in computing and communications. IEEE, 2012.

[17] Frumberg, Fred. "Beyond revival and preservation: contemporary dance in Cambodia." Beyond the Apsara. Routledge India, 2020. 140-154.

[18] Gaston, Anne-Marie, and Tony Gaston. "Dance as a way of being religious." The Oxford Handbook of Religion and the Arts. Oxford University Press, 2014. 182.

[19] Chunduru, Vaishnav, Mrinalkanti Roy, and Rajeevlochana G. Chittawadigi. "Hand tracking in 3d space using mediapipe and pnp method for intuitive control of virtual globe." 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, 2021.

[20] Raghavan, Neethu, and S. Ullas. "Infant movement detection and constant monitoring using wireless sensors." 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2017.

[21] Prasad, Ch Durga, and N. Srinivasu. "Fault detection in transmission lines using instantaneous power with ED based fault index." Procedia Technology 21 (2015): 132-138.

[22] Joseph, Neetha, et al. "A novel approach for group formation in collaborative learning using learner preferences." 2017 international conference on advances in computing, communications and informatics (ICACCI). IEEE, 2017.

[23] Joseph, Neetha, et al. "A novel approach for group formation in collaborative learning using learner preferences." 2017 international conference on advances in computing, communications and informatics (ICACCI). IEEE, 2017.

[24] Pohl, Henning, and Aristotelis Hadjakos. "Dance pattern recognition using dynamic time warping." Sound and Music Computing 2010 (2010).

[25] Jang, Minsu, et al. "Automated dance motion evaluation using dynamic time warping and Laban movement analysis." 2017 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2017.