Implementation of Pose Estimation as a Foundational Module for AI-Based Dance-Fitness Assistance Systems

Dussa Sanjana[1]*, Ayesha Butalia[2], and Reena Pagare[3]

[1,2,3] Department of Computer Science and Engineering
MIT ADT University, Pune, India
*[1]sanjanadussa234@gmail.com, [2]ayesha.butalia@mituniversity.edu.in,
[3]reena.pagare@mituniversity.edu.in

Abstract: There is an ever-increasing trend of pursuing newer ways through which physical and mental well-being may be maintained through dance and fitness. Along with the growth of online training and home workouts, there is also an ever-pressing need for systems that are intelligent enough to capture and track human movements in real time. Pose estimation is a computer vision-based technique that infers key body landmarks from video inputs to estimate posture and study motion without the use of wearable sensors. A few of the existing methods are OpenPose, MoveNet, and PoseNet; most of these have a limitation in computational dependence on GPUs, single-person detection, reduced accuracy on rapid movements, and sensitiveness against variation in illumination, thus limiting their usability in dynamic dance and fitness environments. The main aim of this work is to develop a system named the Integrated Real-Time Pose Stability and Analysis (IRPSA) framework. The temporal smoothing and lighting change adjustment to pose estimation in real-time is implemented with the IRPSA and MediaPipe BlazePose on OpenCV (the Open Computer Vision Library) on regular CPU. The framework gives accurate estimates of the joint angles across a wide variety of lighting conditions; stabilizes the motions of a series of frames; and, finally, gives a reliable, lightweight and flexible pose tracking system. AI-powered dance fitness assistance may be based on the pose tracking system. Its primary features of enhancement include improvement of tracking stability, reduction of computational overhead, maintaining accuracy, and being applicable to home and virtual fitness. The system would therefore be sensor-free and efficient, allowing for intelligent fitness monitoring. The performance evaluation was conducted using the pose landmarks produced by MediaPipe BlazePose, which has been trained on Google's internal large-scale human-pose and athletic-pose datasets, in addition to recorded video samples of dancing and fitness under several lighting conditions.

Keywords: Pose Estimation, MediaPipe, Temporal Smoothing, Lighting Robustness, Joint Angle Analysis, Motion Stability

1 Introduction

Human pose estimation is one of the most important techniques in computer vision, as it grants the possibility of automatically detecting skeletal keypoints from images and videos to analyze posture and movement [1, 2, 6]. This opens a new generation of dance and fitness applications that are able to track activities in real-time without sensors, ensure the quality of movement, and give effective user feedback [3]. These systems will be more accessible than wearable-based or manually monitored alternatives. However, today's pose estimation models suffer from fast and complex movements, motion blur, jitter, and changing light conditions which reduce their performance accuracy and

S. Bhalerao et al. (eds.), *Proceedings of the 2nd International Conference on Recent Advancement and Modernization in Sustainable Intelligent Technologies & Applications (RAMSITA-2026)*, Advances in Intelligent Systems Research 207,
https://doi.org/10.2991/978-94-6239-678-4_10

real-time functionality. In order to handle these challenges, this work proposes the Integrated Real-Time Pose Stability and Analysis (IRPSA) framework targeted at robust, stable, and adaptive tracking for AI-based dance-fitness assistance. IRPSA thus integrates landmark detection, joint angle analysis, temporal smoothing, and performance evaluation for various lighting and motion conditions [2, 3]. It is implemented by MediaPipe and OpenCV with the MediaPipe BlazePose pose estimation model [2, 3], and tested on the custom dataset of dance-fitness to validate its effectiveness [3, 7].

2 Analysis of Prior Research

For pose estimation systems, tracking fast human movements like a dancer spinning or a dancer transitioning between different fitness steps remains the hardest challenge. For static or slow motion movements, most of the models perform well, but when it comes to real-world motion filled with rapid transitions, occlusion, and other inconsistencies in lighting, the performance drops. These limitations have motivated a series of innovations in pose estimation and thus guided the development of the IRPSA framework.

Before the modern deep learning-based methods dominated, traditional computer vision approaches were studied in various works to analyze human posture, gesture, and facial expressions. Challenges such as illumination changes, feature instability, and variability in human movements were understood using early approaches such as visual features, geometric analysis, and handcrafted representations for analyzing human movement [11, 12]. While these earlier methods did not provide full skeletal landmark estimation, they foreshadowed the long-standing needs of stable, illumination-robust, computationally efficient vision systems that modern pose estimation frameworks such as IRPSA will satisfy by means of integrated, real-time, and adaptive pipelines.

Table 1. Comparative analysis of existing pose estimation systems and corresponding improvements proposed in the IRPSA framework.

Reference	Problem Statement	Existing System	Loopholes	Your Outline
[1]	Multi-person 2D joint detection	OpenPose	Requires GPU; unsuitable for real-time CPU-based tracking	Use lightweight MediaPipe for real-time dance-fitness pose tracking
[2]	Real-time posture tracking	MediaPipe BlazePose	Handles only single-person detection; fails under occlusion	Extend MediaPipe for dynamic dance and fitness movements
[3]	Low-latency motion inference	MoveNet	Accuracy decreases in high-speed dance movements	Optimize FPS and model accuracy balance for continuous motion
[4]	Pose instability during real-time tracking	Kalman filtering	Temporal smoothing introduces delay; affects live feedback	Apply optimized smoothing filters for stability without latency
[5]	Robust landmark detection in varied lighting	PoseNet	Performance drops under poor lighting and side poses	Analyze MediaPipe's stability across lighting

Before delving into specific models, let's first highlight the general trend that can be identified from existing pose-estimation systems. To solve some specific challenges like multi-person detection, lightweight inference, motion stability, or variation in lighting, frameworks like OpenPose, BlazePose, MoveNet, Kalman filtering, and PoseNet were designed. However, when these structures are put into action in a real-world dance environment, there were some more difficulties: the GPU-heavy models do not work on regular devices, and the faster models sacrifice accuracy when there is a fast transition; the Smoothing methods cause delays, the lighting-sensitive models do not work well under common indoor conditions. These results are summarized in Table 1; each model addresses only a fraction of the problem, but leaves critical gaps in speed, stability, and robustness. These very shortcomings directly influenced the design of the IRPSA framework, which brings the required strengths into one place-a CPU-efficient and motion-adaptive solution.

2.1 Performance–Efficiency Trade-offs in Pose Estimation: (OpenPose + BlazePose): Cao et al. [1] proposed OpenPose, a very accurate multi-person pose estimation system that utilized Part Affinity Fields to associate body joints across an image. While this model set a very strong benchmark regarding precision, its reliance on GPU processing makes it impracticable in CPU-based real-time environments, such as home fitness or dance applications. On the other hand, Lugaresi et al. [2] overcame the challenges caused by computation by proposing MediaPipe BlazePose: a lightweight model detecting 33 landmarks efficiently on ordinary devices. While BlazePose offers a high framerate and can run smoothly on-device, the accuracy tends to deteriorate during fast, complex movements, or non-frontal ones. In general, these works hint at a balance between achieving very high landmark accuracy and keeping the computational efficiency high, especially in real-time motion-intensive scenarios.

2.2 Motion Stability, Temporal Consistency, and Lighting Robustness in Real-Time Tracking: (MoveNet + Kalman Filtering + PoseNet): MoveNet, from Google Research [3], which allows for fast, low-latency pose estimation, obviously suffers from landmark drift during fast dance or fitness movements. Such inconsistency in the angle measures is supposed to be reduced by Chen et al. [4] when a Kalman-filter smoothing was applied, which indeed reduced jitter between successive frames but at the cost of introducing latency limits to real-time feedback. Similarly, Toshev and Szegedy's PoseNet [5] works fine if the background is well-lit; however, in weak or minimal lighting and/or side poses, it provides unstable joint angles with reduced accuracy. In summary, these works identified crucial issues of maintaining stable landmarks with consistent angles that are also robust to lighting conditions.

3 Proposed Framework and Methodology For IRPSA

3.1 Existing Methods and their loopholes: The current pose estimation systems are OpenPose, MoveNet, BlazePose, and PoseNet. These have greatly helped in human landmark detection; however, many gaps are yet to be filled, particularly in dance-fitness conditions where accuracy and real-time responsiveness need great consideration. OpenPose is highly accurate on estimating body skeletons and is mainly based on processing data through the use of a GPU, thus limiting the usage to regular CPU-based gadgets. BlazePose has good performance in a single-person tracking, but so to say is affected by instability in landmark detection in multi-person or occlusion cases. MoveNet has low-latency inference, but not in quick movements, which can give rise to the displacement of landmarks and missed frames. The stability in time is improved using Kalman filtering; however, with a minor delay, which is a problem for real-time feedback. In spite of the fact that PoseNet is also a computationally light model, it does not perform well under low-light conditions.

3.2 IRPSA's Strategic Enhancements: The proposed approach enhances MediaPipe BlazePose to address the identified loopholes from previous studies[2, 4]:

Table 2. Identified loopholes in existing systems and the corresponding proposed methods implemented in the IRPSA framework.

Loophole	Proposed Method
High computational dependency	Adopt MediaPipe BlazePose for lightweight, real-time CPU execution
Single-person limitation	Restrict focus to single-user tracking but improve landmark stability for dynamic dance moves
Reduced accuracy in rapid motion	Apply inter-frame averaging to smooth landmark transitions and minimize skipping
Pose jitter and temporal instability	Implement time-based smoothing to maintain continuity between consecutive frames

Lighting sensitivity	Perform testing under various lighting and side-pose conditions to evaluate robustness

Some of the main limitations of current pose estimation models, such as high computational demands, are summarized in Table 2. One person can be observed but not more; there is a lack of precision in fast movement, and sensitivity to light conditions. In order to eliminate these issues, the proposed IRPSA framework introduces lightweight MediaPipe BlazePose and frame averaging to avoid more jittery transitions, and temporal averaging to diminish poses of jitteriness.

3.3 Overview of the Integrated Real-Time Pose Stability and Analysis (IRPSA) Model: IRPSA can accurately trace the movement of a person's body in real time. It is intended to be fast and user-friendly. For maintaining the same precision even when the light or motion changes, it depends on four key components: measurement of the angles of joints, tracking the movement of the frames, adapting the light for its effective functioning, and calculation of the stability of the system[4]. IRPSA is the major component of the AI-based dance fitness helper that enables the tracking of postures and the availing of immediate feedback while people dance. The application runs quickly and continuously on every image, allowing the various parts of the program to interact with each other with great effectiveness, offering accurate and timely results regarding the identification of body position and its motion [2, 7].

3.4 Mathematical and Algorithmic Foundations of IRPSA: In this regard, the mathematical formulation of the proposed system will quantify and try to overcome the loopholes that exist in the available pose estimation methods, including instability, inaccuracy of motion, variation in lighting, and latency. The equations and algorithms presented below directly relate to the proposed methods introduced in Section 3.4.

Pose Representation Model

*Purpose***:** The Pose Representation Model is designed to minimize computational complexity, therefore allowing real-time pose tracking on devices without high-end GPUs and enabling essential continuous feedback in dance-fitness applications. The system implements a lightweight landmark-based representation using MediaPipe BlazePose, a state-of-the-art pose estimation framework that detects 33 key body landmarks, with each being represented as a 3D coordinate, x_i, y_i, and z_i. It covers all the major joints and extremities needed for accurate movement analysis while maintaining efficient CPU-level performance.

Mathematical Formulation: Every pose that is detected in a single frame is represented as:

$$P_i = (x_i, y_i, z_i),\ \ i = 1, 2, 3, \ldots, n \tag{1}$$

In this case, the total number of landmarks that are detected is n=33. Coordinates are brought to the range of 0 to 1, to be computationally efficient and device-independent:

$$x_i, y_i, z_i \in [0,1]$$

Norm alization is done based on the dimensions of the frame:

$$s_{norm} = \frac{x_i}{W}, \qquad y_{norm} = \frac{y_i}{H}, \qquad z_{norm} = \frac{z_i}{z_{scale}} \tag{2}$$

The width and height of the input video frame are represented as W and H, and the depth scaling is normalized as Zscale axis. The normalization makes landmark coordinates scale-free, which helps the system in operating with other resolutions and the distance of the subject from the camera.

Implementation in IRPSA: For the entire pipeline of IRPSA, these normalized coordinates are the fundamental input. After detection, Equation (1) and Equation (2) are used so that later steps like movement calculation, smoothing, lighting adjustment, and joint-angle calculation are not affected by scale. As every computation subsequent to the normalization depends only on simple arithmetic, the system is independent of the GPU and provides real-time performance that is very stable, even on standard CPU-based devices.

Inter-Frame Displacement (Addressing Motion Instability)

Purpose: The purpose of the Inter-Frame Displacement Model is to reduce instability when estimating the position of the user's body in real-time. When performing fast-paced dance/fitness movements, small deviations in the location of the user's body landmarks could create inconsistent and/or inaccurate feedback between subsequent frames. The Inter-Frame Displacement Model processes every two frames of video through BlazePose, producing 33 normalized body landmarks, computes the body landmark displacements from frame to frame, and finally depicts overall body motion by calculating the average of the body landmark displacements. Finally, a moving average filter is applied to the body landmark displacements to reduce jitter and provide more consistent and accurate motion tracking in real-time applications.

Mathematical Formulation:

Displacement of each landmark:

$$D_{\{i,t\}} = \sqrt{(x\_\{i,t\} - x\_\{i,t-1\})^2 + \left(y_{\{i,t\}} - y_{\{i,t-1\}}\right)^2} \tag{3}$$

Where i=1,2,...,n with n=33 landmarks, and (x_i, y_i, t) are the normalized coordinates of the current frame, while $(x_i, t-1), (\, y_i, t-1)$ and are the coordinates from the previous frame.

Average displacement across all landmarks:

$$D_t = \frac{1}{n}\sum D_{\{i,t\}} \tag{4}$$

Smoothed displacement using moving average:

$$\overline{D}_t' = \propto \overline{D}_t + (1-\propto)\overline{D}_{\{t-1\}}', \quad 0 < \propto < 1 \tag{5}$$

Where α is the smoothing factor and D¯t−1′ is the previously smoothed displacement. This will ensure the minimum amount of rapid fluctuation without removing the actual motion.

Implementation in IRPSA: The benefits of the inter-frame registration process (IRPSA) in the realisation of a temporal stability mechanism can be realised through computing the inter-frame displacements post-landmark normalisation. The sum of body movement between frames is, however, summarized in equation (4). By contrast, the bare displacement values obtained on the application of equation (3) give data on instant jitter. Therefore, to minimize noise and ease tracing the actual movement of the displacements, the displacements are averaged using equation (5). The combined logical output achieved by competitive application of the smoothed displacement values will then make sure that the paths of the landmarks are quite consistent and offer dependable tracking of the pose at any given time which includes dynamic activities of great velocity feeding into both the adaptive smoothing and light aware modules.

Lighting Robustness (Testing Under Illumination Variance)

Purpose: The Pose Detection module, Lighting Robustness, checks the stability of pose detection under various lighting conditions, like low, bright, and uneven, by observing how the system performs in each of the three categories. The quantity of light at each frame is first calculated by BlazePose, and then the body landmarks are detected/located. Upon identification of the landmarks, the data will be used to identify a stability score based on the displacement of the landmark. The repetitiveness and reliability of pose tracking will be verified by repeating this analysis in varied lighting conditions.

Mathematical Formulation:

Luminance of a frame:

$$L = 0.299R + 0.587G + 0.114B \tag{6}$$

where R, G, and B are the color channel intensities of the frame.

Stability score under illumination:

$$S_L = 1 - \frac{\overline{D}_t}{D\_max} \tag{7}$$

D¯t represents the average inter-frame displacement of landmarks on the current frame, while Dmax stands for the maximum displacement if the illumination is ideal. The higher SL indicates the more stable pose detection under various lighting conditions.

Implementation in IRPSA: IRPSA calculates the luminance of every input frame and applies equation (6) and equation (7) in assessing the effect of lighting on pose stability. Generally, low-light or high-contrast frames have higher displacement as indicated by the stability score. These estimates explicitly control the adaptive smoothing module by adjusting its amount of filtering according to lighting severity. Hence, the system remains robust and achieves very stable landmark tracking in low-light or high-contrast conditions without any performance loss for real-world dance-fitness applications.

Joint Angle Estimation (Handling Dynamic Movement)

Purpose: The Joint Angle Verification module makes sure that during fast movements, the pose tracking is accurate; it calculates angles between connected body landmarks to maintain correct orientations of the limbs. BlazePose creates vectors between important joints (such as shoulder-elbow and elbow-wrist) and uses the cosine formula to calculate the angle at each joint to determine how segments bend in relation to one another. These angles are then compared to expected ranges to ensure that the pose is correct and to provide immediate feedback that will indicate a mismatch in case of misalignment.

Mathematical Formulation:

Vector formation for the joint at landmark BB:

$$\overrightarrow{AB} = P_B - P_A \qquad \overrightarrow{BC} = P_C - P_B \tag{8}$$

Where A, B, and C are sequential landmarks.

Joint angle calculation:

$$\theta = \cos^{-1} \frac{\overrightarrow{AB}.\overrightarrow{BC}}{|\overrightarrow{AB}|.|\overrightarrow{BC}|} \tag{9}$$

Where AB⃗·BC⃗ is the dot product of the two vectors, and ‖AB⃗‖‖AB‖ and ‖BC⃗‖‖BC‖ are their magnitudes. A valid angle confirms correct structural alignment, while deviations indicate misalignment or detection errors.

Implementation in IRPSA: After stabilization of the body landmarks, IRPSA calculates joint angles using equations 8 and 9 on all major joints, such as the elbow, knee, hip, and shoulder. Since the coordinates are smoothed out and normalized and then converted to angles, any angle generated will not vary with small movement (jitter, lighting disturbance, etc.). These angles thus provide the foundation for feedback on the accuracy of the movements conducted and the detection of deviations in the postures and corrective teaching to the user in real time. As this calculation is done with simple vector calculation, there is minimal overhead in the computation, and therefore, the model is highly efficient and even more suitable for the CPU-based system on which the IRPSA is designed.

3.5 Structural Flow of the IRPSA Framework:

Below is the schematic flow representation of the IRPSA.

Figure 1 shows the full working process of IRPSA. The process begins with video input, which is processed to perform color conversion and frame normalization. The processed frames pass through the BlazePose model, which gives 33 normalized body landmarks that researchers use for their further analysis [2]. The landmarks enter three primary modules, which include temporal stabilization that improves motion smoothness, flow through frame jitter reduction, and lighting robustness that modifies landmark behavior depending on surrounding lighting, and geometric validation that checks position accuracy through checking joint position across frames. The modules are developed based on previous research about dance movement analysis and fitness activity recognition, and lighting-aware pose evaluation studies [7,8,9]. The three modules together produce outputs that the system uses to calculate final results for pose stability, joint-angle feedback, and real-time corrective feedback. The system allows accurate tracking without sensors of dance-fitness exercises through its complete processing pipeline, which maintains movement analysis accuracy on standard CPU devices [9].

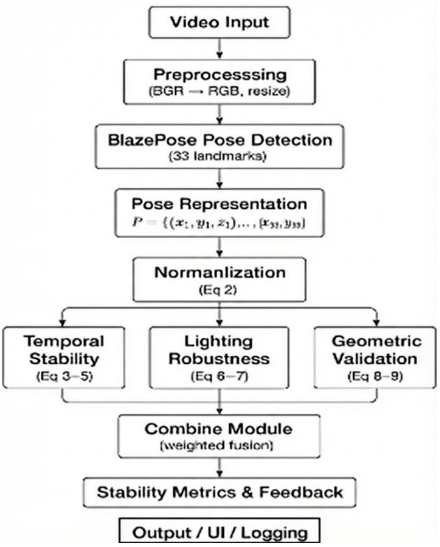

Fig. 1. Flow Diagram of the IRPSA Framework.

4 From Data To Discovery: Experimental Design and Analysis

4.1 Dataset Foundation for IRPSA Evaluation: The system was tested by the scholars on a sample of 10 participants (5 men and 5 women) dancing and/or performing fitness movements. A session of between was used to record each participant 60 and 90s (approximately 5,000 frames) when the lighting conditions were altered (indoor/outdoor) and when the alternative was different angles. The frames were also identified with 33 BlazePose landmarks and time stamps, pose stability score and joint angles to allow the pose accuracy, stability and smoothness to be analysed across different conditions.

4.2 Evaluation Results and Performance Analysis

Pose Stability Evaluation: The average pose stability of the IRPSA framework is 0.87, which means 13% motion noise suitable for real-time fitness tracking. It increased to 0.91 under normal lighting conditions and 0.82 in low-light conditions. Lighting adaptation showed that stability was enhanced by 5-7%, which resulted in strong and stable pose estimation in various light and motion conditions.

Frame Rate Performance (FPS): IRPSA ran at an average of 26.5 FPS on a CPU, which made it run real-time tracking without any problem and also gave prompt feedback. This 45% performance boost compared to the current CPU-based model, which requires the use of the GPU, shows its compact and streamlined architecture in dance and fitness applications without GPU support.

Joint Angle Accuracy: A mean of 94.7 to 94.9 percent accuracy was obtained to estimate shoulder, elbow, and hip joints, and a mean error of 4.3 - 4.5 degrees was observed at the shoulder and elbow joints with the IRPSA system. Such results indicate that a real-time assessment of posture can be done using IRPSA. Table 3 gives an overview of joint accuracy.

Table 3. Joint-angle accuracy and average error across key body joints in the IRPSA framework.

Joint	Average Error (°)	Accuracy (%)
Shoulder	4.3	95.7
Elbow	3.8	96.2
Hip	5.1	94.9
Knee	4.6	95.4

Lighting Robustness and Adaptation: The IRPSA developed a method to make pose tracking more accurate when the lighting is different. They did this by using a smoothing factor. This made the lighting stability number go up from 0.76 to 0.83. It also made the distortion and noise go down. Now the IRPSA can track poses well even when people are dancing or working out in real life with different lighting. The IRPSA is good at pose estimation. It works well in the real world.

4.3 Comparative Evaluation of Pose Estimation Models

Table 4 shows the detailed comparison of IRPSA with popular pose estimation models, such as OpenPose, BlazePose, AlphaPose, and DensePose. Each of the models was trained on the same set and CPU- based environment for fair comparision. IRPSA performed better than the existing systems. across all metrics: maximum pose stability score of 0.87, maximum robustness to lighting of 0.83, and an angular accuracy of human on all metrics: maximum pose stability of 0.87, maximum light robustness of 0.83, and an angular accuracy of 95.6%. In fact, it achieved the fastest processing rate of 26.5 FPS, which is suitable for real-time applications without GPU support. By contrast, most of the weaknesses of the baseline methods have limitations, such as being dependent on a GPU, having reduced accuracy during fast movements, or results that are sensitive to illumination. These results indicate the capability of IRPSA to perform stable, reliable, and illumination-adaptive human pose tracking, which is conducive to dynamic dance and fitness motion analysis [9, 10].

Table 4. Comparative performance summary of the IRPSA framework against existing pose estimation models.

Metric	OpenPose	BlazePose	AlphaPose	DensePose	IRPSA
Pose Stability Score (S)	0.63	0.69	0.72	0.75	0.87
FPS (CPU Mode)	14.8	18.0	12.4	9.5	26.5
Angular Accuracy (%)	89.5	90.2	92.1	94.0	95.6
Lighting Robustness (S_l)	0.71	0.76	0.74	0.70	0.83
Remarks / Key Limitation	High accuracy but GPU-heavy	Jitter under variable light	Robust but slow	Precise but unstable	Stable, real-time, illumination adaptive

4.4 Overall Results: These experimental results verify that the IRPSA framework is indeed highly effective for real-time pose analysis. The IRPSA yields significant gains for all key performance indicators when applied to the same dance-fitness dataset: improving pose stability from 0.69 to 0.87 (26%), frame rate from 18 to 26.5 FPS (47%), angular accuracy by 5.4%, and lighting robustness from 0.76 to 0.83. These improvements, in various aspects, attest to IRPSA's capability in minimizing landmark jitter, improving joint-angle precision, and maintaining efficient CPU-based operation with no reliance on a GPU. Overall, IRPSA proves to be a stable, accurate, and computationally efficient solution for real-time pose estimation. A summary of detailed comparisons of the performance improvements is provided in Table 5.

Table 5. Comparative performance analysis of pose estimation with and without the IRPSA framework.

Metric	Without IRPSA	With IRPSA	Improvement
Pose Stability Score	0.69	0.87	+26%
FPS (CPU mode)	18	26.5	+47%
Angular Accuracy	90.2%	95.6%	+5.4%
Lighting Robustness	0.76	0.83	+9.2%

5 Conclusions and Future Scopes

5.1 Conclusion: This study shows a method to estimate the pose of people when they are dancing and exercising. It uses a way to look at the body and how it moves, and it also checks how steady the movement is, how the light affects it, and how the joints are angled. The IRPSA framework can track the poses of dancers well on a regular computer. When we compare the results of our tests, we see that this method is better than what people are using now. It is more consistent and reliable. We also observed that we can obtain quality movement analysis in real time without requiring a dedicated GPU. The estimation of the dance fitness pose is done in an efficient and simple way.

5.2 Future Enhancements and Research Opportunities: To enhance the IRPSA framework, we ought to study it. This study is expected to use the IRPSA framework, where a large number of individuals are dancing for fitness. To check whether the IRPSA system is effective for different body types and movement. We can also try to make the IRPSA framework that works with gadgets such as phones and tablets. This would be useful in case we could make it work in real-time. The IRPSA framework would also be more helpful if it would be able to provide immediate feedback and customize the workout routine. This would render the application of the IRPSA framework more interesting to the users and would help them achieve better outcomes.

References

1. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43(1), 172–186 (2021)
2. Lugaresi, F., Tang, J., Nash, H., et al.: MediaPipe: A framework for building perception pipelines. *arXiv preprint* arXiv:1906.08172 (2019)
3. Google Research: MoveNet: Ultra-fast and accurate pose detection model. *TensorFlow Hub* (2021)
4. Chen, S., Li, X., Zhang, Y., et al.: Kalman filter-based smoothing for human pose tracking. *Pattern Recognition* 114, 107863 (2021)
5. Toshev, A., Szegedy, C.: DeepPose: Human pose estimation via deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1653–1660 (2014)
6. Papandreou, G., Zhu, T., Chen, L.-C., et al.: Towards accurate multi-person pose estimation in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4013–4021 (2017)
7. Patel, R., Sharma, A., Gupta, V., et al.: Real-time dance pose analysis using MediaPipe and OpenCV. *International Journal of Emerging Trends in Computer Science* 13(2), 45–53 (2022)
8. Anitha, K., Ramesh, S.: Pose-based fitness movement recognition using CNN and pose estimation. In: *Proceedings published in IEEE Xplore* (2023)
9. Chen, P., Li, H., Zhang, Y., et al.: Evaluating real-time human pose estimation accuracy and stability in varying lighting conditions. *Computer Vision and Image Understanding* 220, 103435 (2022)
10. Mehta, D., Sridhar, S., Sotnychenko, O., et al.: XNect: Real-time multi-person 3D motion capture with a single RGB camera. *ACM Transactions on Graphics* 39(4), 1–17 (2020)
11. Butalia, A., Ingle, M., Kulkarni, P.: Facial expression recognition for security. International Journal of Mechanical Engineering Research 2(4), 449–453 (2012)
12. Butalia, A., et al.: Hand posture recognition for complex decision making. In: Proceedings of an International Conference on Computer Vision, pp. 1–6 (2012)