ENHANCED HUMAN POSE ESTIMATION UNDER OCCLUSION USING GRAPH-BASED OPTIMIZATION TECHNIQUES: A STUDY WITH G2O.

Supervised by
Mohammad Mahadi Hassan
Associate Professor
Dept. of CSE, IIUC

Co Supervised by
Md. Khaliluzzaman
Associate Professor
Dept. of CSE, IIUC

Presented by
Shirajul Islam Shakur (C213040)
Alfaz Uddin Emon(C213057)
Md. Mohaiminul Islam(C213067)

Contents

Research Gap

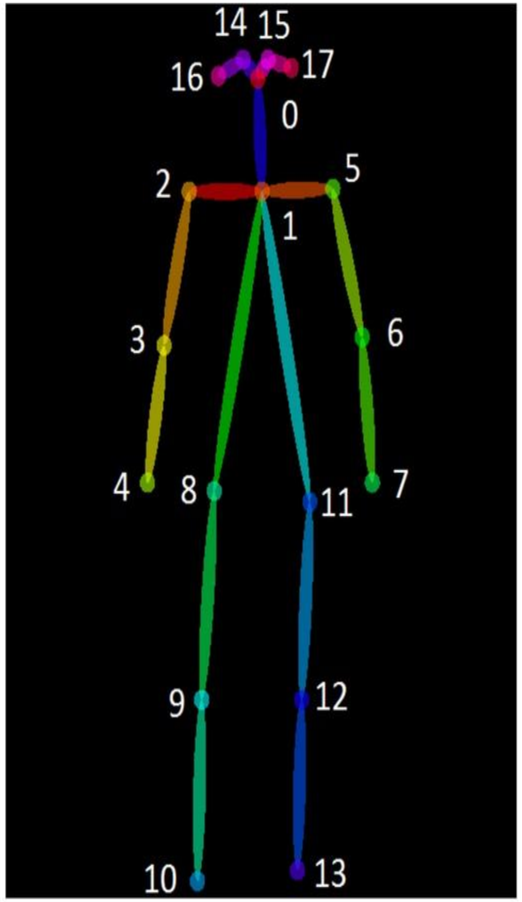
- Current models struggle with accurate keypoint prediction under occlusion.
- High computational costs, especially while processing video dataset.

Motivation

- Pose estimation is crucial for applications in healthcare, sports, physical therapy, virtual reality, and more.
- We use deep learning models like OpenPose, which struggle with occlusion or high computational costs.
- Graph-based methods like G2O improve accuracy by handling occlusion and reducing computational load.

Objectives

- Optimize keypoint predictions using graph-based method to handle occlusion.
- Use G2O-based optimization to enforce anatomical and temporal constraints for stable pose tracking.
- Develop methods that maintain accuracy even when parts of the body are occluded or when the input data is noisy or blurred. complex



Challenges

- **Occlusion:** Occluded frames make it difficult to achieve accurate pose estimation.
- **Temporal Consistency:** Maintaining smooth and logical transitions for keypoints between consecutive video frames is challenging, especially with occlusion
- **Dataset Limitations:** This study faced a lack of a proper, pre-existing video dataset tailored to the specific occlusion scenarios being investigated.

Related Works

g2o: A General Framework for Graph Optimization :

The g2o framework is a general, open-source C++ library for efficiently solving graph-based nonlinear least squares problems, widely used in SLAM and bundle adjustment

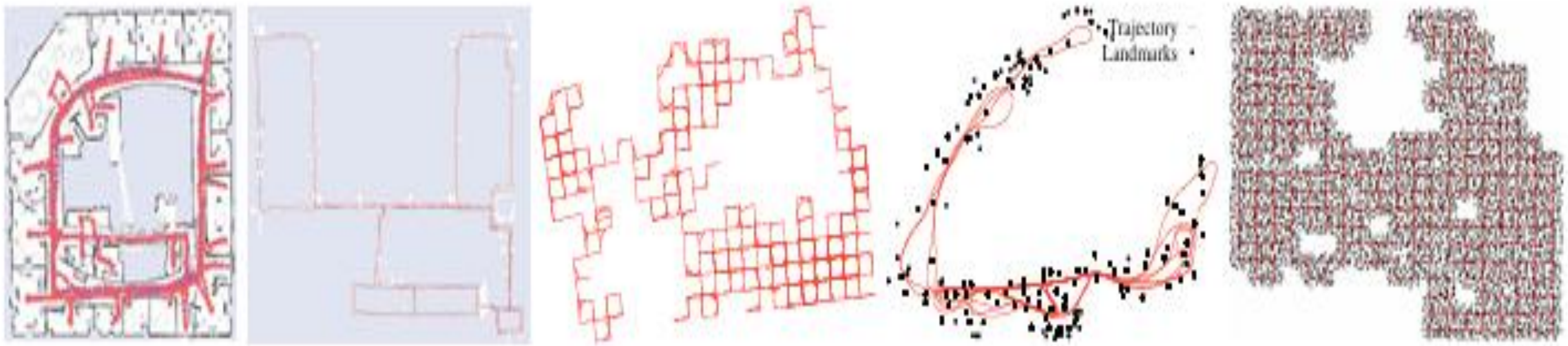


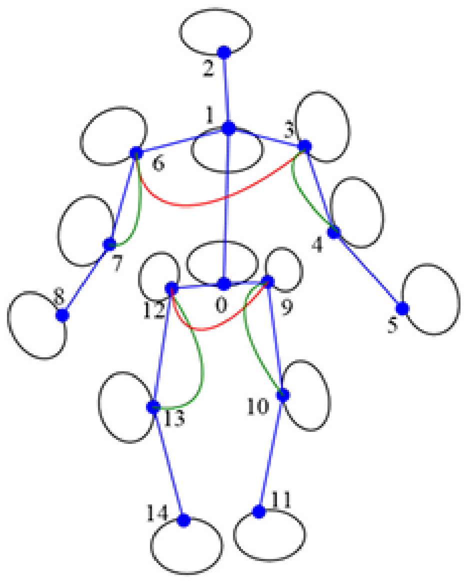
Fig. 6. 2D Datasets used for evaluating $\mathbf{g^2o}$. From left to right: 2D pose-graph of the Intel Research Lab; 2D pose-graph of the Killian Court; Manhattan3500, a simulated pose-graph; 2D dataset with landmarks of the Victoria Park; and Grid5000, a simulated landmark dataset.

Related Works Contd.

Real Time Monocular 3D Human Pose Estimation Based On G20:

G2O-Pose uses general graph optimization (g2o) to achieve real-time monocular 3D human pose estimation with improved accuracy and stability.

Limitation: Its performance decreases under severe occlusions and relies heavily on good 2D pose initialization.

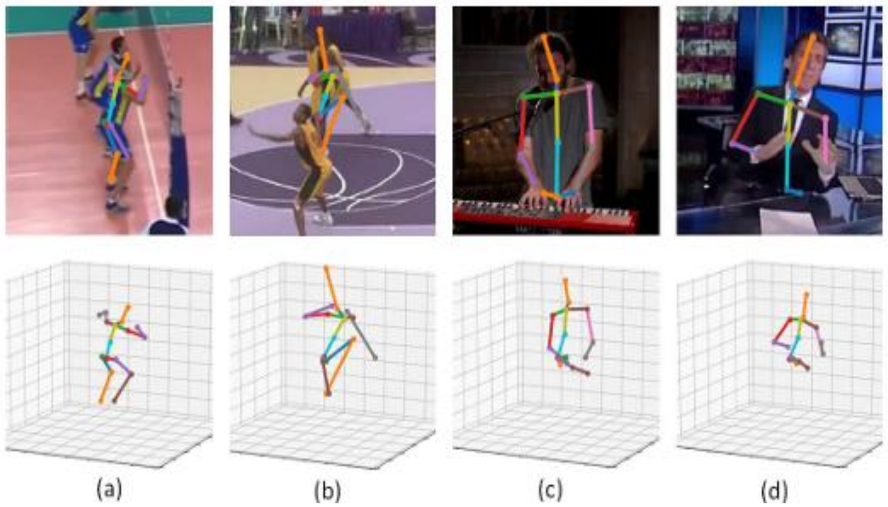


Related Works Contd.

Occlusion-aware human pose estimation using graph-based optimization:

This paper proposes an occlusion-aware human pose estimation method that leverages graph-based optimization to refine poses under challenging conditions.

Limitation: The approach is computationally intensive and may struggle with real-time performance on complex multi-person scenes.



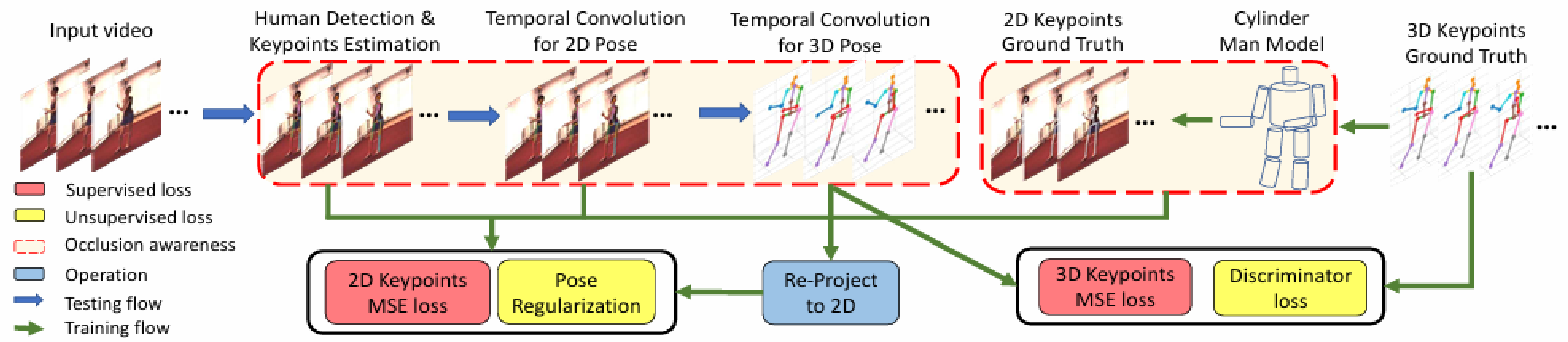


Figure 2. The framework of our approach, best viewed in color.

Baseline Framework: OpenPose

- A real-time framework for detecting 2D human keypoints for multiple people in an image or video.
- It was used in this research as the initial step to extract 2D keypoint data from all video frames.

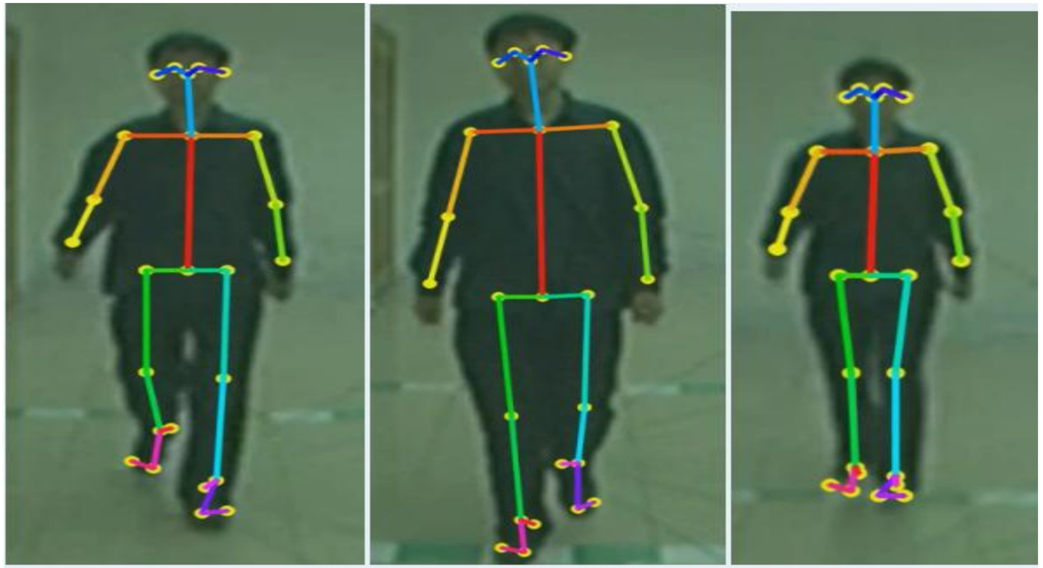


Figure 4.2.1: Non-Occluded Keypoint Detection

Baseline Framework: OpenPose

Performance and Limitations

- **Clear Views**: OpenPose accurately detects human poses when the body is fully visible
- **Under Occlusion**: Its performance drops significantly when parts of the body are hidden.
- **Reason**: It lacks knowledge of human anatomy, leading to inaccurate poses and inconsistent bone lengths when visual information is limited.

Optimization Process

- **Levenberg-Marquardt Algorithm**: This method is a combination of **Gauss-Newton** and **gradient descent**, which adjusts the variables in each iteration to find the optimal solution.
- **Gauss-Newton Method**: A widely used method for solving non-linear least squares problems that approximates the solution by iterating to improve the variables estimates.
- **Gradient Descent**: In cases where the problem is convex or close to convex, gradient descent is used to minimize the error function.

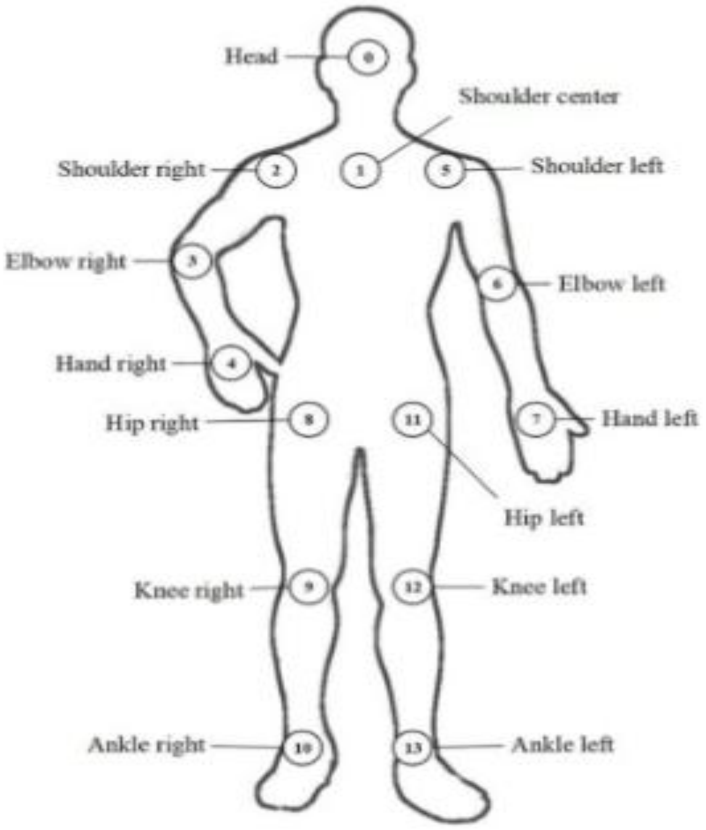
Theoretical background

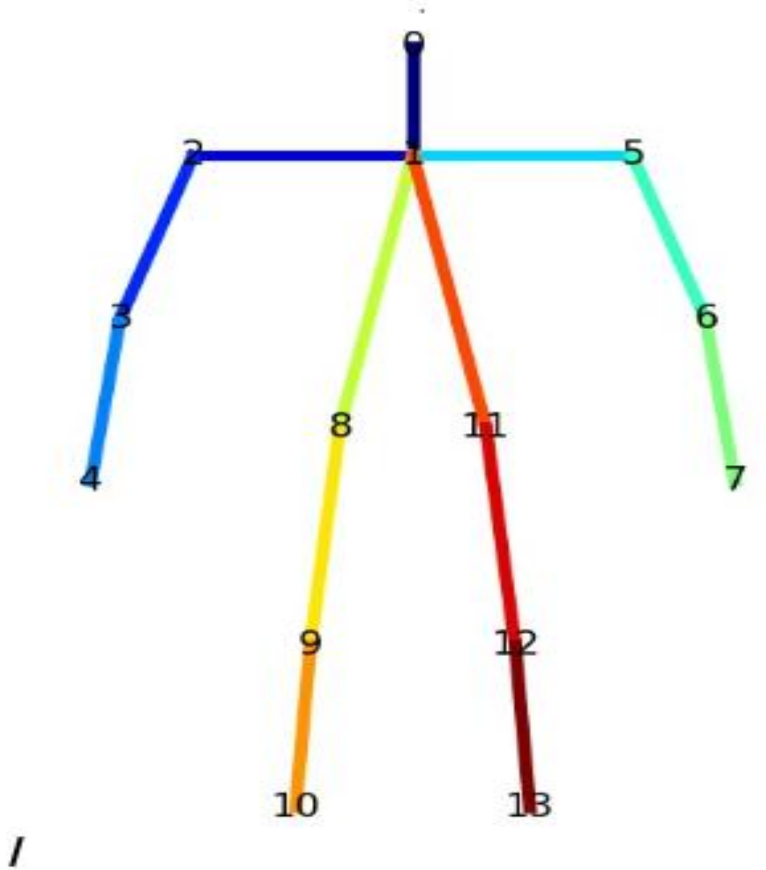
Graph Representation: In the context of pose estimation, a graph is used where:

- **Vertices:** Vertices are the representation of **keypoints** in pose estimation.
- **Edges** represent the constraints or measurements between these vertices, such as bone connections

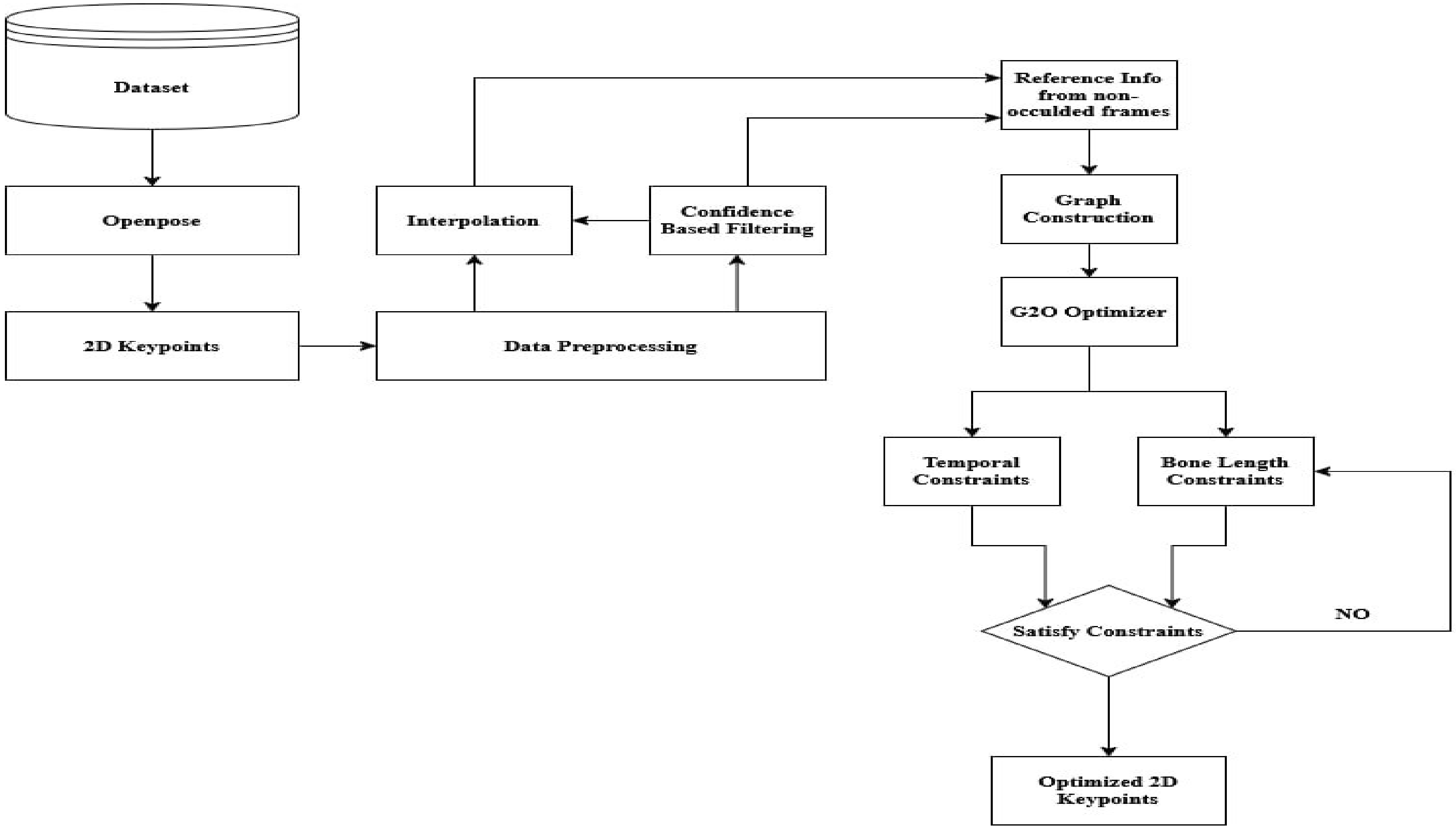
G2O: G2O (General Graph Optimization) is an open-source C++ system designed for solving non-linear least squares problems, which are common in robotics and computer vision. It is applied here to optimize 2D human keypoints by enforcing temporal and anatomical constraints.

Overview of body key joints



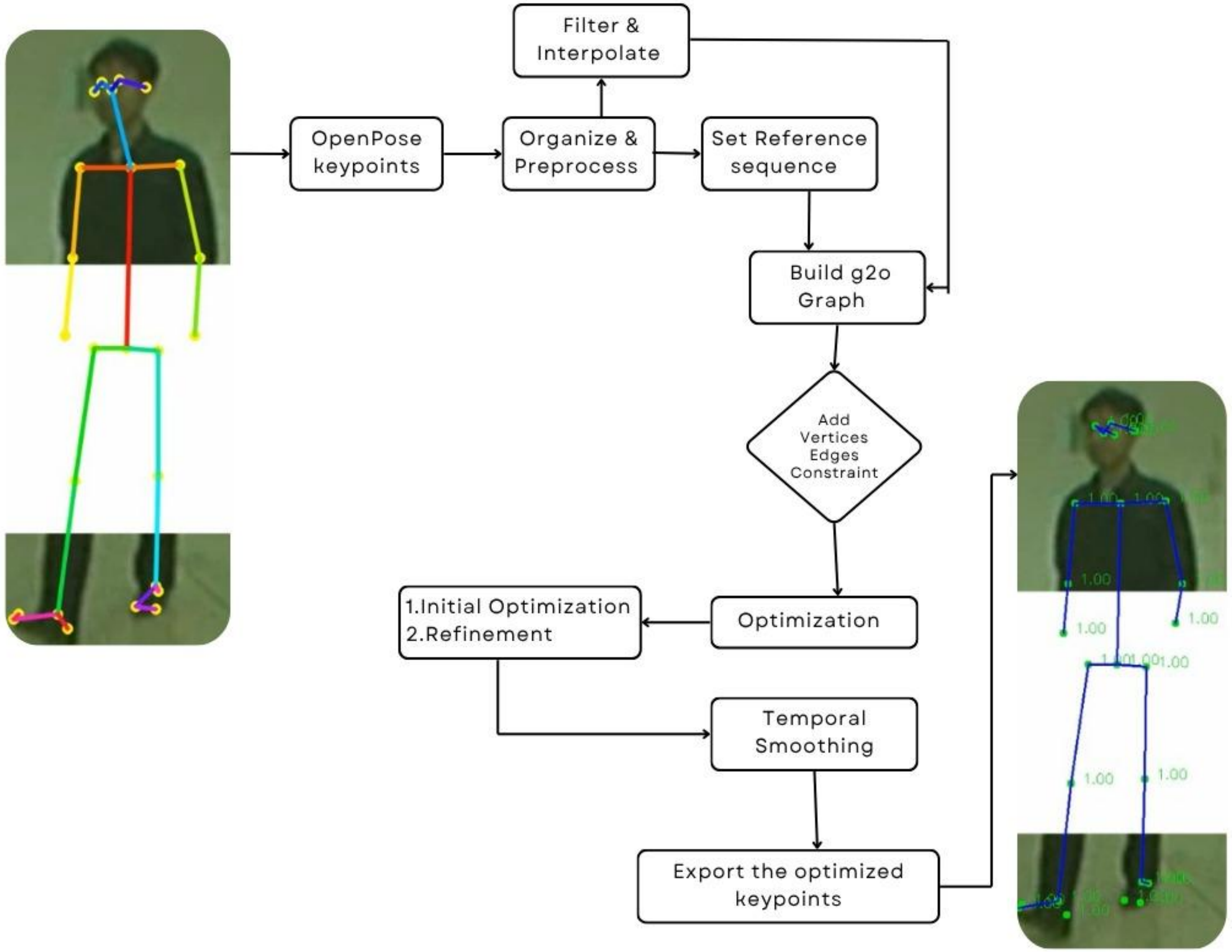


Proposed Methodology



Overview of AT G20 Framework

we introduce ATG2O (Anatomical and Tem poral Graph Optimization for Obstacle-Affected Pose Estima tion), a novel algorithm designed to refine 2D human pose estimations by integrating anatomical and temporal constraints within a graph optimization framework.



Experimental Settings

- **Programming Language: Programming Language**: Python was used for the implementation, leveraging a few important libraries.
- **Keypoint Estimation**: **OpenPose** was used for initial 2D keypoint detection for each frame of the video.
- **Libraries Used**:
  - **G2O**: For graph-based optimization to refine the pose keypoints.
  - **NumPy**: For numerical operations, matrix handling, and data manipulation.
  - **JSON**: For parsing and handling keypoint data stored in JSON format.
- **Environment**: The code was executed on Jupyter Notebook.

Dataset

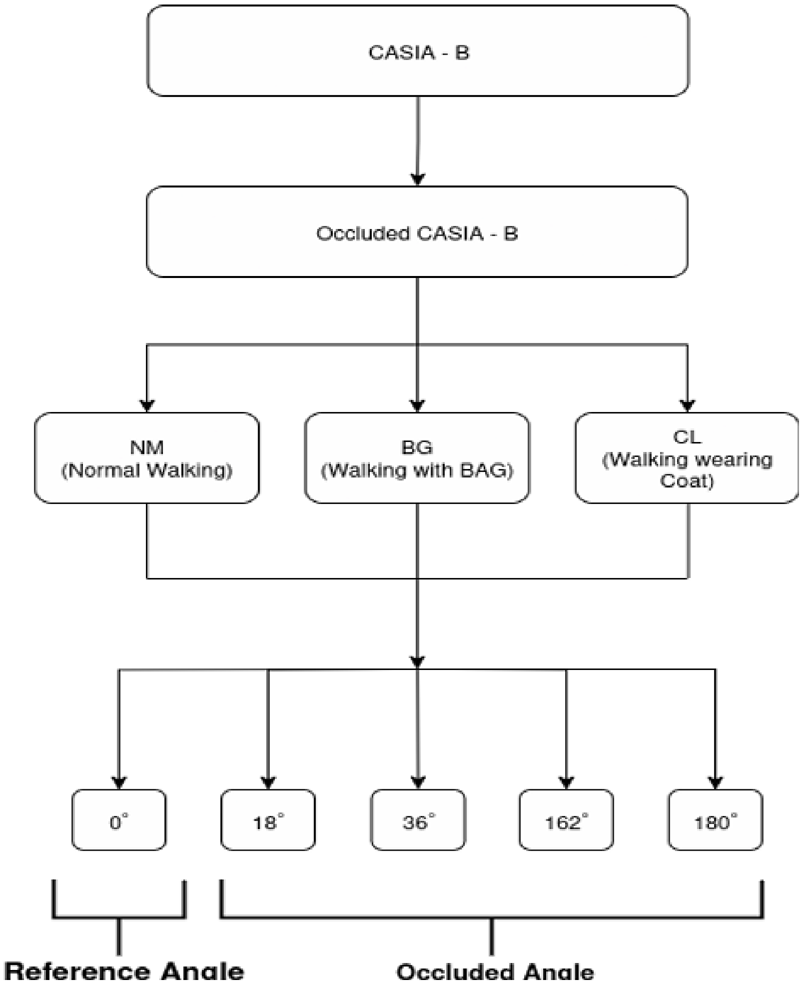
- The used OccCASIA-B dataset in this study consists of video modify by us, captured from three different camera angles or scenarios.
- Two of the angles/scenarios include occlusions for 15–25 frames in each sequence, caused by external objects.

CASIA-B

OccCASIA-B

Dataset Structure

In our OccCASIA-B construction, these three walking conditions are preserved in separate folders named NM, BG, and CL.



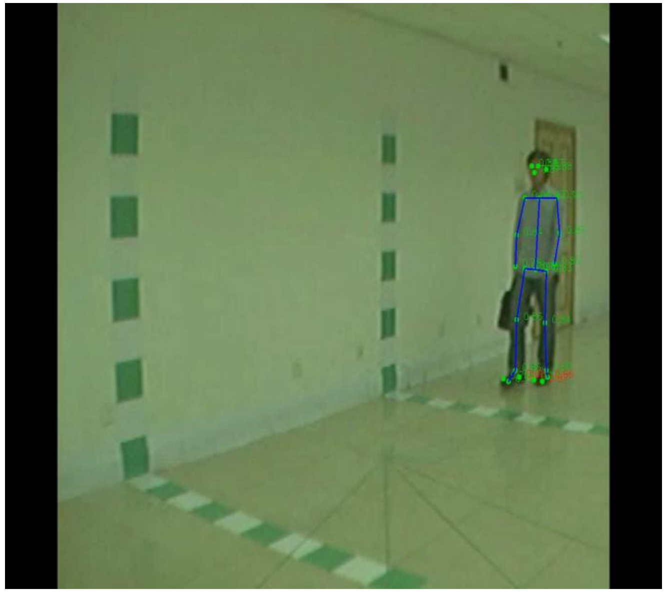
Openpose Detection

OpenPose is able to perfectly detect the pose when there is no occlusion

Walking while carrying a bag   Walking while wearing a coat   Normal walking

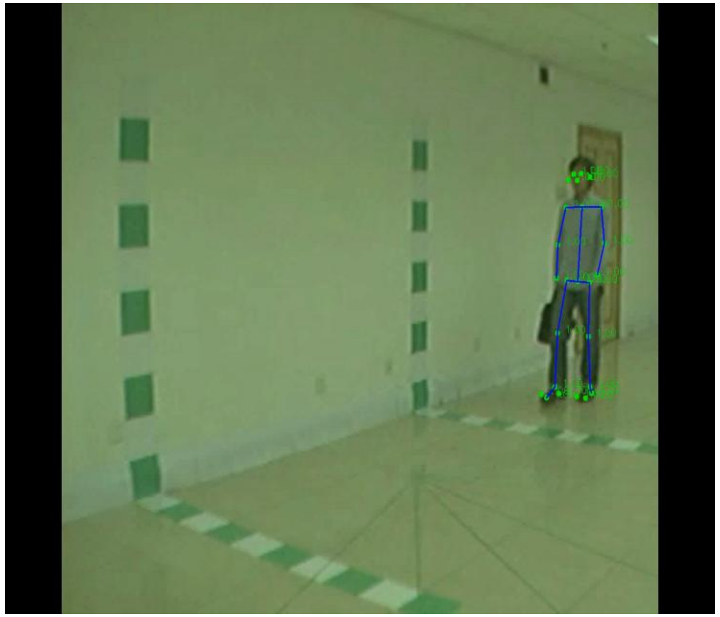
Openpose

These videos were then processed in OpenPose to get joint keypoints from our participants. Here we demonstrate the OpenPose output while coming under occlusion.



G2O Optimization

These videos were optimized in G2O to get joint more optimized keypoints from openpose keypoints.



Comparative Analysis

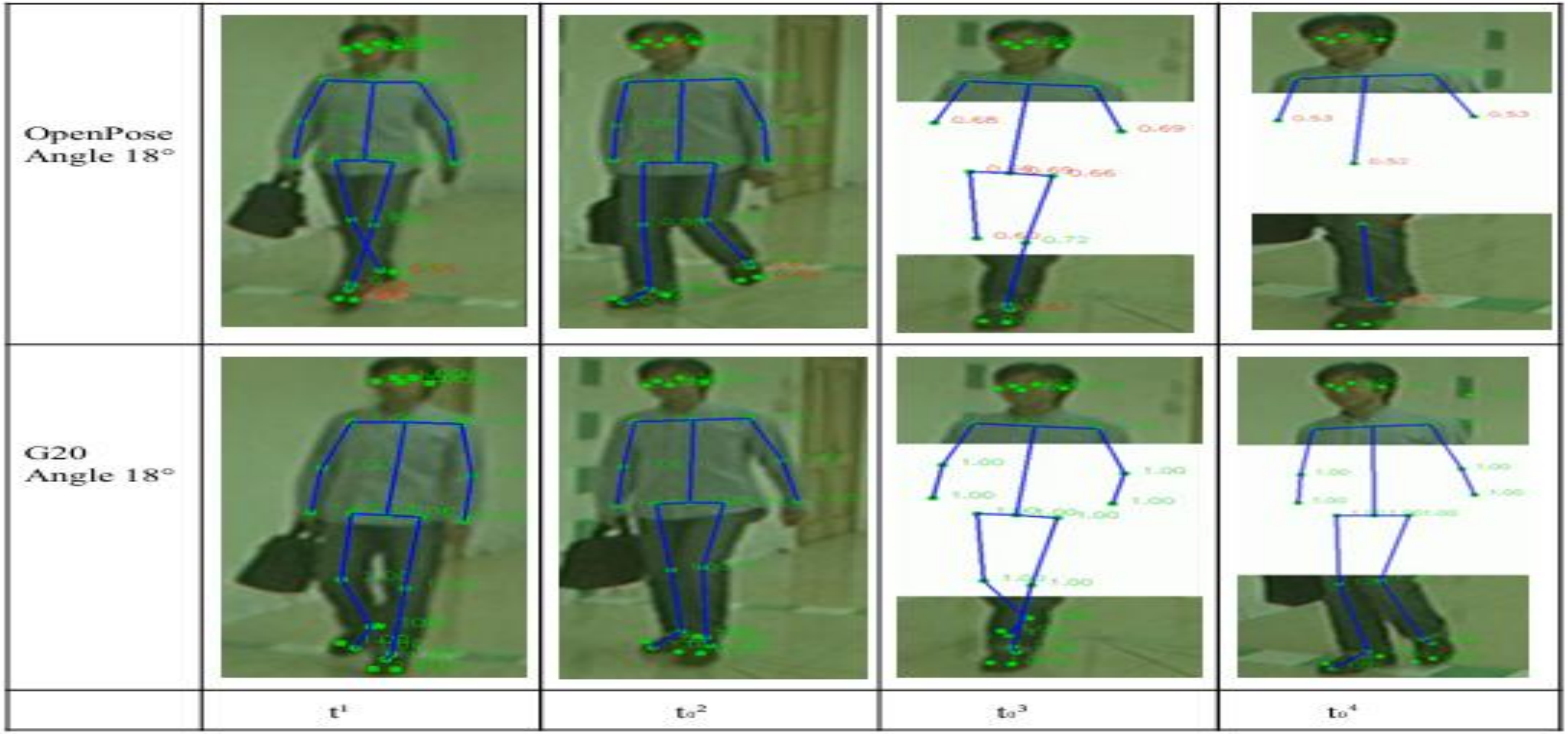


Table 4.2.1 : walking with bag _ 18° Openpose VS G2O

Comparative Analysis Contd.

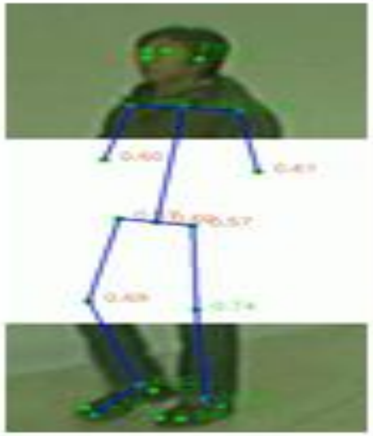
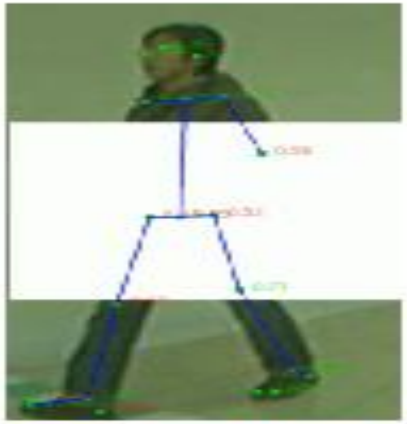
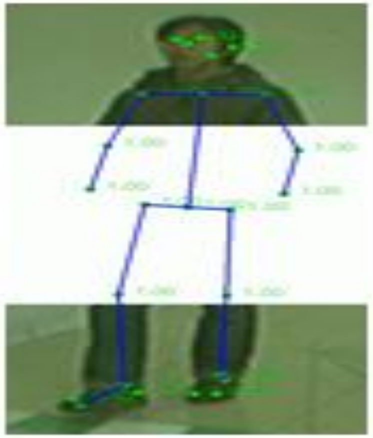
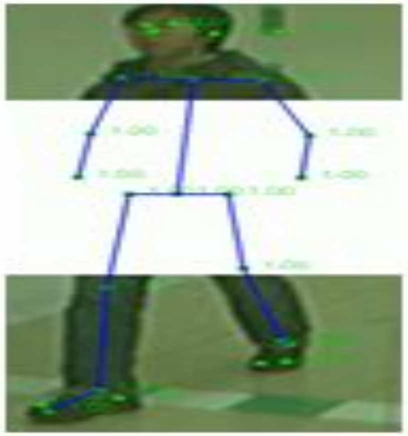
OpenPose Angle 36°				
G20 Angle 36°				
	t^1	t_0^2	t_0^3	t_0^4

Table 4.2.5 : walking wearing a coat 36° Openpose VS G20

Comparative Analysis Contd.

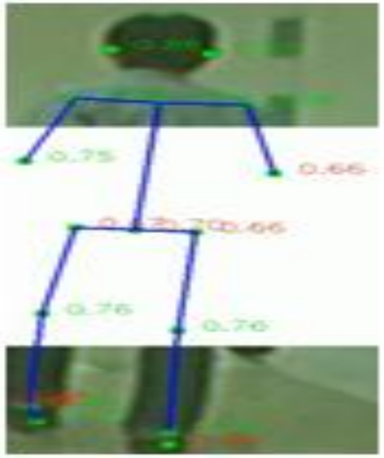
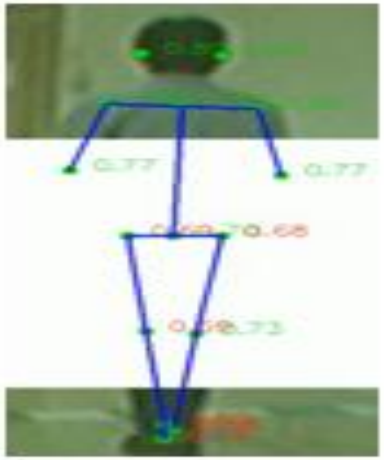
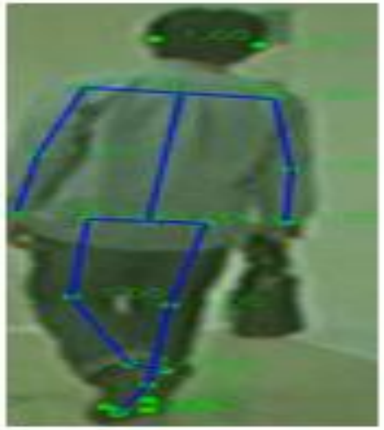
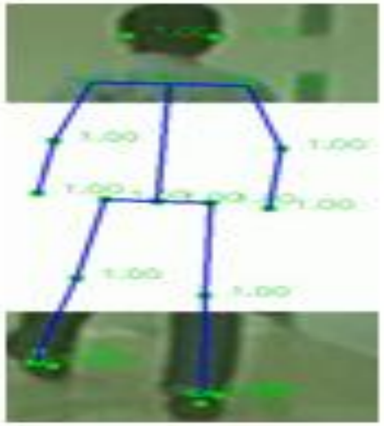
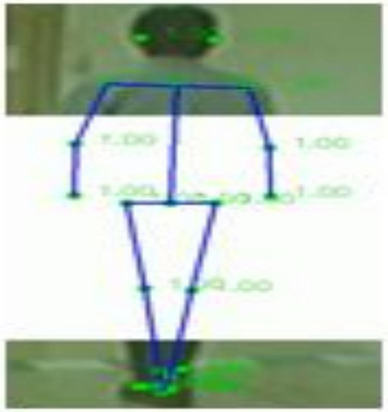
OpenPose Angle 162°				
G20 Angle 162°				
	t^1	t_0^2	t_0^3	t_0^4

Table 4.2.3: Walking with bag 162° Openpose VS G2O

Comparative Analysis Contd.

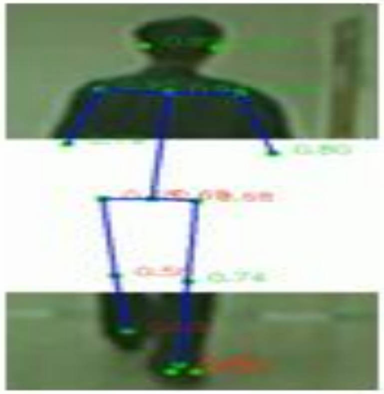
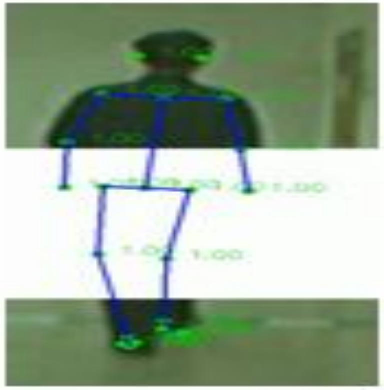
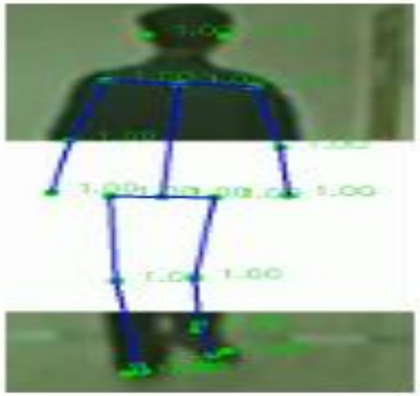
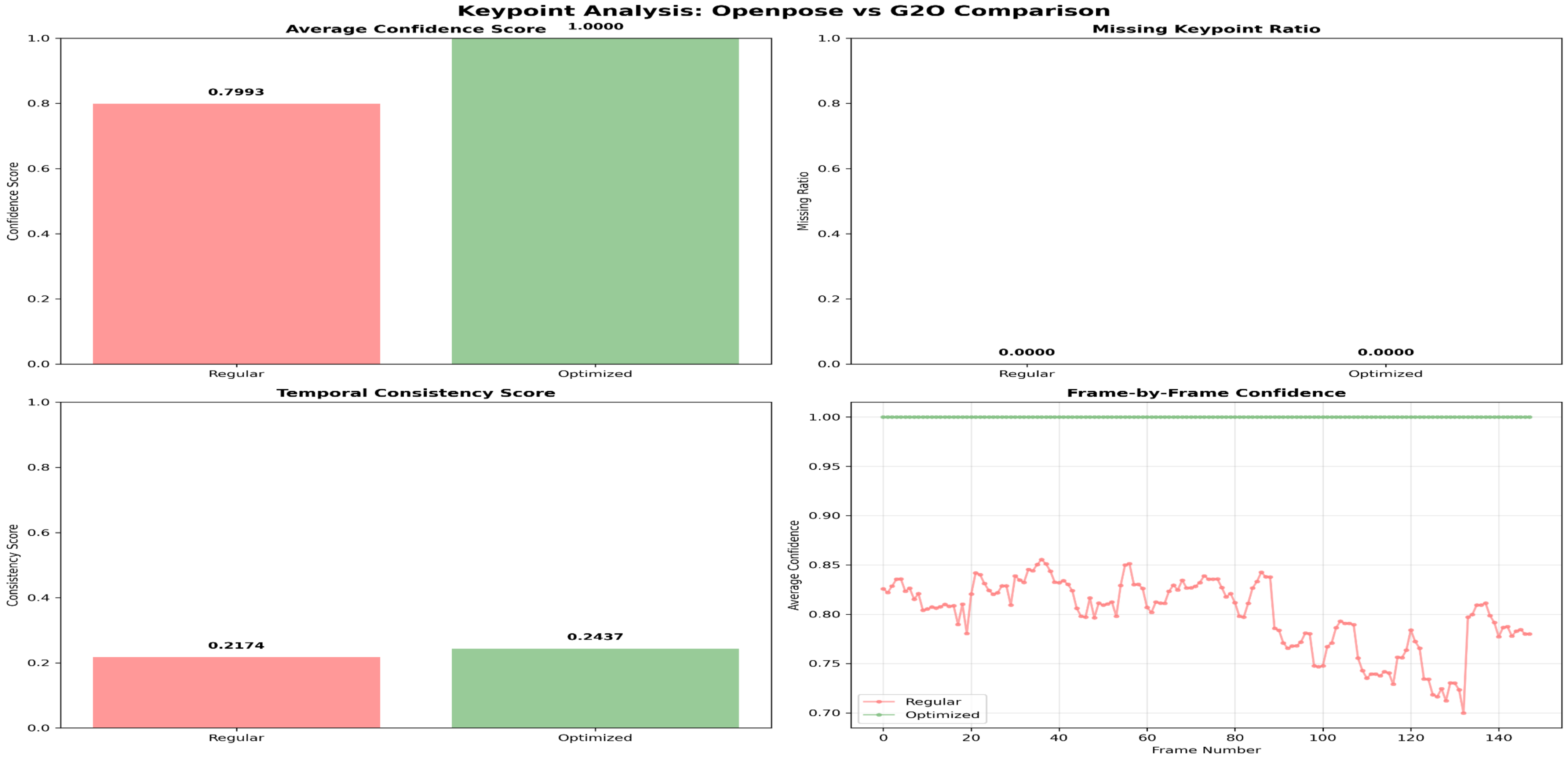
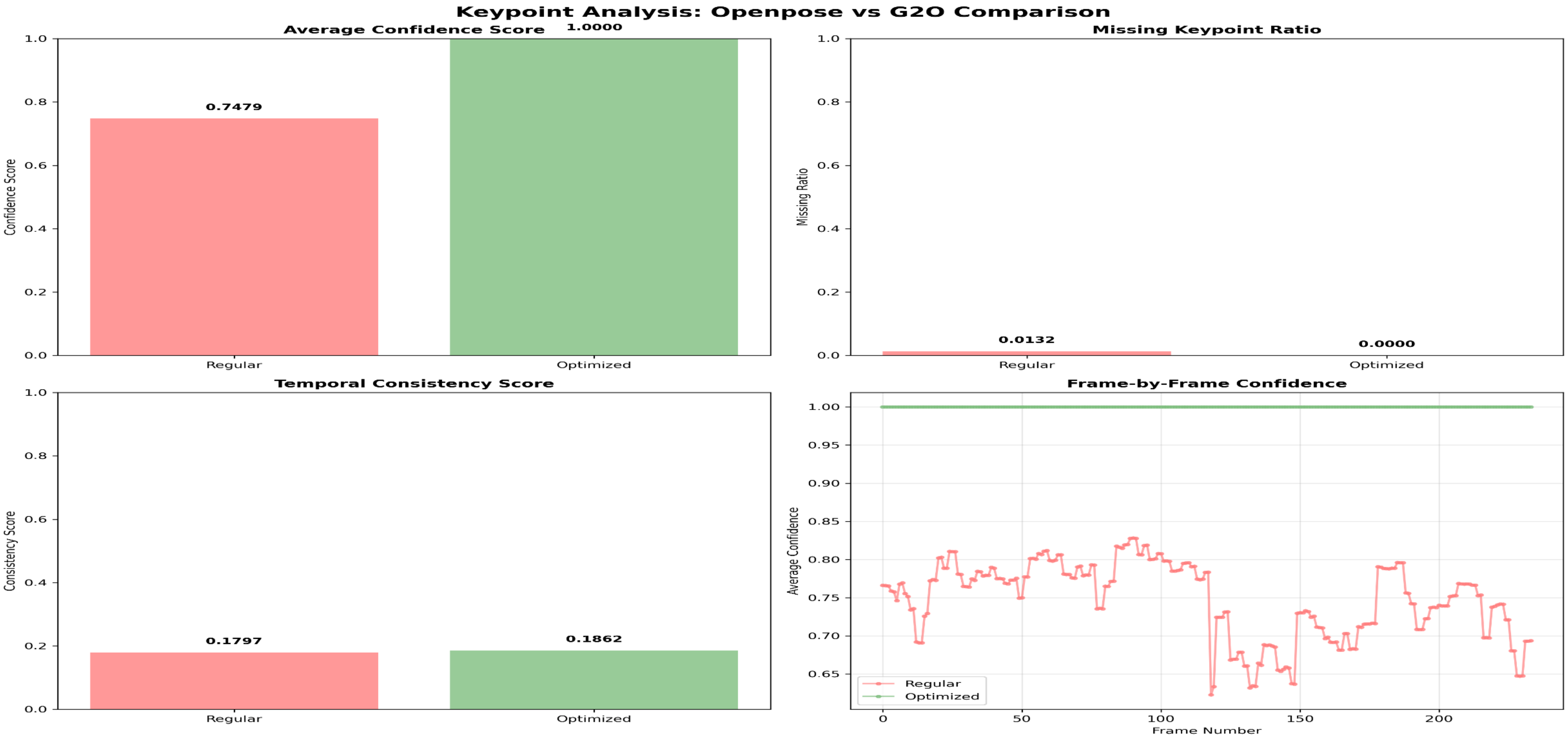
OpenPose Angle 180°				
G20 Angle 180°				
	t^1	t_0^2	t_0^3	t_0^4

Table 4.2.7 : Normal walking 180° Openpose VS G2O

Experimental Analysis



Experimental Analysis



Conclusion

- **G2O Framework**: G2O outperforms OpenPose in pose estimation under occlusion, improving bone length consistency and temporal smoothing. It can recognize the facial expression in the diverse datasets

- **Anatomical & Temporal Constraints**: The integration of constraints refines keypoint predictions and ensures anatomical fidelity in occluded frames.

Future Work

- Expanding the dataset to include a wider range of activities, more participants, and diverse occlusion patterns
- Come up with more well defined constraints after learning them from a large dataset, which was a problem in our case.

References

- [1] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, Jan. 2021.
- [2] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” International Conference on Learning Representations (ICLR), 2021.
- [3] X. Zhou, M. Zhu, Z. Deng, and X. Yuan, “G2O-based optimization for monocular human pose estimation,” IEEE Transactions on Image Processing, vol. 31, pp. 1235- 1248, 2022.
- [4] F. Wang, M. Cheng, X. Liu, and J. Luo, “Occlusion-aware human pose estimation using graph-based optimization,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 401–412.
- [5] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, “G2O: A general framework for graph optimization,” Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3607–3613, 2011.
- [6] J. Martinez, R. Hossain, J. Romero, and J. J. Little, “A simple yet effective baseline for 3D human pose estimation,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2659–2668.
- [7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
- [8] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, “2D human pose estimation: New benchmark and state of the art analysis,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3686- 3693.
- [9] pavllo, D., Feichtenhofer, C., Grangier, D., & Auli, M. (2019). “3D human pose estimation in video with temporal convolutions and semi-supervised training.” CVPR 2019.