International Islamic University Chittagong

Department of Computer Science and Engineering

A Thesis Proposal on:

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

Submitted By:

Name: Shirajul Islam Shakur

ID: C213040

Name: Alfaz Uddin Emon

ID: C213057

Name: Md. Mohaiminul Islam

ID: C213067

Approval of the Supervisor

Mohammad Mahadi Hassan

Associate professor

Approval of the Co-Supervisor

Md. Khaliluzzaman

Associate Professor

Table of Contents

1. Introduction …….....1
2. Motivation……..1
3. Literature Review……...2
4. Problem Statement…….5
5. Aims and Objectives ……..……5
6. Methodology……..6
7. Application…….7
8. Expected Contribution……7
9. Expected Outcomes.……...7
10. Work Plan & Distribution……...8
11. Conclusion……..8
12. References……..9

1. Introduction

Human pose estimation, the task of identifying and localizing key body joints from images or videos. It is a very demanding topic in computer vision for its wide range of applications such as healthcare, sports analytics, motion analysis and virtual and augmented reality. However, one of the key challenges when it comes to human pose estimation is the problem occlusion. Despite advances in deep learning-based methods, achieving highly accurate and efficient pose estimation remains challenging, particularly in real-world scenarios involving occlusions, complex poses, and varying camera angles. Traditional optimization methods often face difficulties in balancing accuracy with computational efficiency, especially when dealing with complex spatial dependencies and noise in the data.

This study proposes a novel approach utilizing General Graph Optimization (G2O) to refine human pose estimation by treating human keypoints as a kinematic graph. By leveraging spatial and temporal constraints, this optimization framework ensures anatomically plausible and occlusion-robust pose estimation, improving real-world applicability.

2. Motivation

The motivation behind this work is the increased demand in many movement-based industries for accurate human pose estimation. While current pose estimation methods struggle with occlusion handling and computational efficiency. Traditional methods rely on neural network-based architectures which also proves to be extremely costly. And thus, reducing that cost by relying on other methods is also an important aspect of our work. So, we decided to utilize optimization frameworks such as G2O, or General Graph Optimization which can handle nonlinear optimizations with efficiency, making it suitable for modeling complex relationships of keypoints and applying constraints to ensure biomechanical acceptability. Graph-based optimization applies constraints on bone length and joint angles, making the estimated poses closer to the natural movement of a human. In this way, with the integrated constraints, a pose estimation model will more precisely infer missing keypoints, resulting in smooth and reliable pose sequences even under occlusion conditions.

3. Literature Review

SN	Title	Author	Dataset	Methodology	Findings
1	Occlusion-Aware Networks for 3D Human Pose Estimation in Video **[1]**	Yu Cheng, Bo Yang, Bo Wang, Wending Yan, and Robby T. Tan1	Human3.6M Dataset: This is a large-scale 3D human pose dataset containing 3.6 million images	The methodology employs 2D confidence heatmaps and optical-flow consistency to filter out unreliable key points. It utilizes 2D and 3D temporal convolutional networks (TCNs) to enforce temporal smoothness and reconstruct occluded key points. Additionally, the Cylinder Man Model generates occlusion-labeled training data and applies pose regularization constraints.	The proposed occlusion-aware framework improves 3D human pose estimation by effectively handling occlusions, reducing estimation errors, and outperformi ng state-of-the-art methods.
2	g2o: A General Framework for Graph Optimization **[2]**	Rainer Kummerle, Giorgio Grisetti , Hauke Strasdat ,Kurt Konolige,Wolfr am Burgard	Manhattan3500 Dataset, Victoria Park Dataset, Venice Dataset, Scale-Drift Dataset	g2o optimizes graph-based nonlinear error functions by leveraging sparse connectivity, advanced solvers, and efficient numerical techniques for SLAM and Bundle Adjustment.	The g2o framework efficiently optimizes graph-based nonlinear error functions, outperformi ng many problem-specific algorithms in SLAM and Bundle Adjustment.
3	A simple yet effective baseline for 3d human pose estimation **[3]**	Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little	HumanEva, Human3.6M- This is currently the largest publicly available dataset for 3D human pose estimation, consisting of 3.6 million images.,	A lightweight feedforward network with residual connections, batch normalization, and dropout maps normalized 2D joint locations to 3D poses in camera coordinates.	A simple deep network lifting 2D joint inputs to 3D significantly reduces error, revealing

			MPII Dataset		that 2D detection quality largely drives 3D pose estimation performance.
4	G2O-Pose: Real-Time Monocular 3D Human Pose Estimation Based on General Graph Optimization **[4]**	Haixun Sun, Yanyan Zhang ,Yijie Zheng, Jianxin Luo * and Zhisong Pan	Training and Testing: Typically, subjects S1, S5, S6, S7, and S8 are used for training, while subjects S9 and S11 are reserved for testing.Depth Considerations:	The study presents G2O-Pose, a real-time monocular 3D human pose estimation method using general graph optimization with multiple constraints, including bone proportion recovery, human orientation classification, and reverse joint correction.	G2O-Pose significantly improves running speed while maintaining comparable accuracy to traditional methods, achieving 32 FPS on consumer hardware and outperforming non-deep learning approaches when depth variation is minimal.
5	OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields **[5]**	ZheCao, Gines Hidalgo, Tomas Simon, Shih-EnWei, and YaserSheikh	The study introduces a publicly released annotated foot dataset, which is used to evaluate foot keypoint detection results,MPII for training and testing,	The methodology involves a bottom-up approach where a feedforward network predicts 2D confidence maps for body part locations and vector fields for PAFs, which encode the association between body parts. The system iteratively refines PAF predictions and confidence maps to output accurate 2D key points for all individuals in the image	The study demonstrates that combining body and foot estimation into a single model enhances accuracy and reduces inference time, while the use of Part Affinity Fields (PAFs)

					significantly improves runtime performance and accuracy in multi-person 2D pose estimation.
6	DSPose: Dual-Space-Driven Keypoint Topology Modeling for Human Pose Estimation **[6]**	Anran Zhao, Jingli Li, Hongtao Zeng, Hongren Cheng and Liangshan Dong	The experiments were conducted using the MPII dataset, which is commonly used for evaluating human pose estimation models	The methodology involves a Transformer-based feature extraction for accurate keypoint features, integration of physical space correlations to address localization errors, and the use of a graph convolutional neural network (GCN) to fuse spatial and feature correlations for improved pose estimation	The proposed dual-space-driven topology model, DSPose, achieved competitive results in human pose estimation tasks, outperforming several existing methods on the MPII dataset by achieving four best and four second-best indicators across various metrics

4. Problem Statement

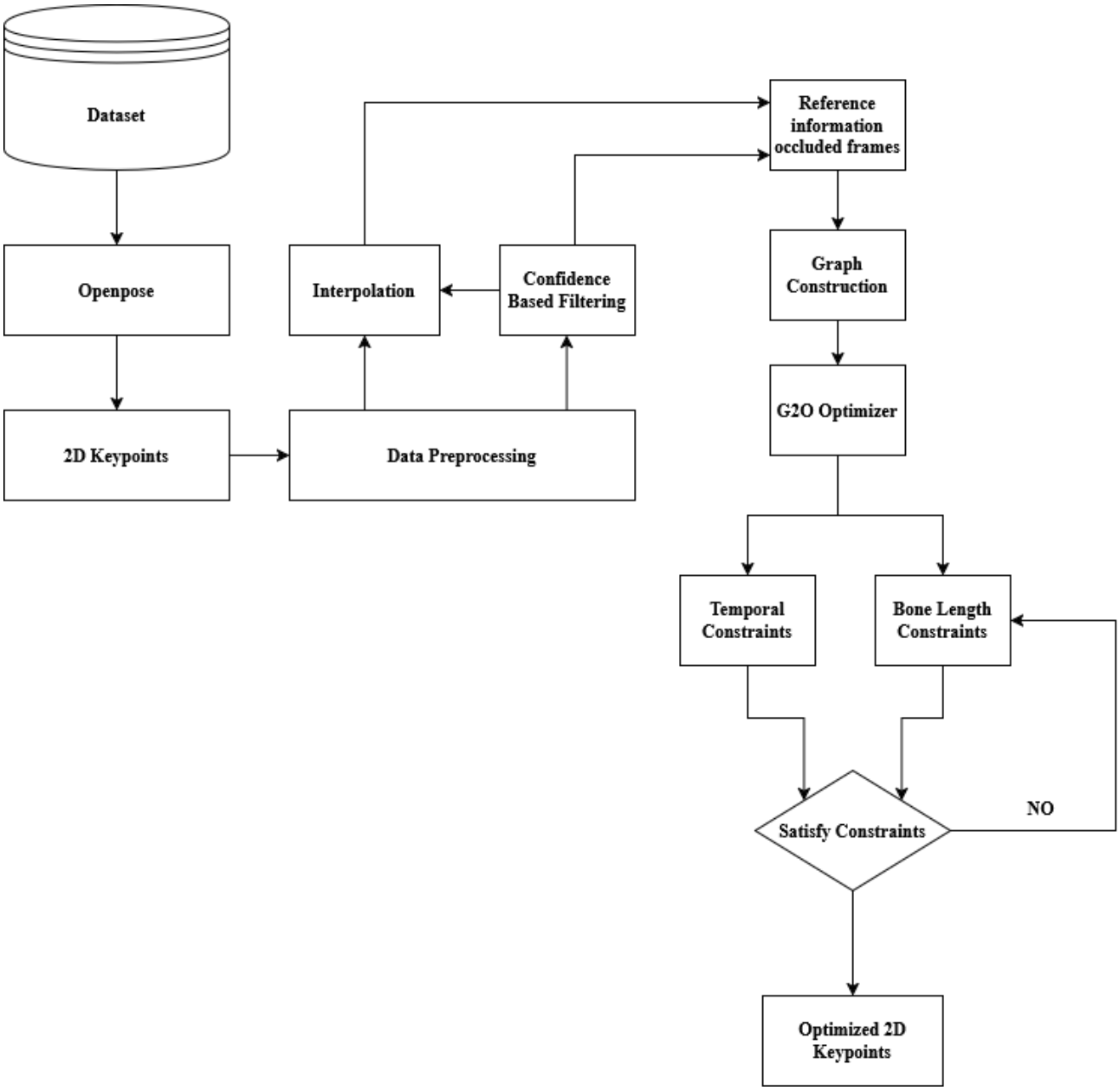
Human pose estimation under conditions of partial or full occlusion presents significant challenges, particularly in fields such as healthcare, sports, and surveillance. While deep learning models perform well in controlled environments, their accuracy declines when body parts are obscured, which can negatively impact medical diagnosis, patient monitoring, and movement analysis. Traditional models rely solely on visual cues and lack mechanisms to infer missing keypoints or maintain movement continuity under occlusion, often resulting in noisy, inconsistent, or anatomically implausible pose estimations. To address these limitations, there is a critical need for methodologies that integrate detected keypoints with contextual information and biomechanical constraints to infer missing data, thereby improving reliability in real-world applications. This work has a significant social impact by enhancing healthcare through improved patient monitoring, optimizing sports training, and increasing safety in surveillance systems. Ethically, it ensures data privacy by using anonymized human data and adhering to strict ethical guidelines. Additionally, from an environmental perspective, our approach prioritizes sustainability by implementing energy-efficient algorithms that reduce computational costs and lower overall energy consumption, making the system both effective and responsible.

5. Aims and Objectives

There are some of the key aims objectives of this research are as follows:

- To utilize the framework General Graph Optimization for human pose estimation based on optimizing the skeleton key points.
- Develop methods that maintain accuracy even when parts of the body are occluded or when the input data is noisy or blurred.
- Implement and refine optimization algorithms within the G2O framework that effectively handle the complex graph structures of keypoints and constraints.

6. Methodology



The proposed method starts with extracting 2D keypoints from a dataset using OpenPose. Since some keypoints may be missing or inaccurate, interpolation and confidence-based filtering are applied to improve the data. The keypoints then go through preprocessing to refine them before optimization.

Next, a graph is constructed using reference information from frames where keypoints are visible. This helps establish relationships between keypoints over time. The G2O optimizer then improves the keypoints by applying temporal constraints (to ensure smooth movement) and bone length constraints (to maintain realistic body proportions).

The optimization process continues until all constraints are satisfied. The final output is a set of optimized 2D keypoints, which are more accurate and reliable for pose estimation.

7. Application:

G2O is widely used in various fields for optimizing variables over a graph structure:

- SLAM (Simultaneous Localization and Mapping.
- Pose Graph Optimization.
- Human Pose Estimation.
- Object Tracking.
- AR/VR.

8. Expected Contributions:

• G2O improves the precision of keypoint localization by optimizing over a graph-based representation of human pose. This allows for more accurate placement of keypoints by refining initial estimates and incorporating constraints that ensure anatomically plausible configurations.

• G2O can optimize the computation involved in pose estimation by efficiently solving large scale optimization problems. This can lead to faster processing times and reduced computational resources compared to traditional methods, making real-time applications more feasible.

9. Expected Outcomes:

• The expected outcomes of a thesis focusing on General Graph Optimization (G2O) for localizing keypoints in human pose estimation would generally include advancements and contributions in several key areas.

• Enhanced precision in detecting and localizing human keypoints, leading to more accurate pose estimation.

• Increased robustness of the pose estimation system to occlusions, noise, and missing data by effectively using the graph-based optimization to infer missing or occluded keypoints.

• Enhanced ability of the system to generalize to a wide range of poses, body types, and environmental conditions.

10. Work Plan & Work Distribution

Work Plan

Month	March				April				May			
Week	1	2	3	4	5	6	7	8	9	10	11	12
Theoritical Knowledge Gain												
Literature Review and Research Planning												
Data Collection and Preparation												
Methodology Development												
Initial Model Training and Testing												
Analysis and Implementation												
Writing the Thesis Report												
Revision and Preparation for Pre-Defense												
Finalization and Submission												

Work Distribution

Team Member	Responsibilities
Alfaz Uddin Emon	Lead the literature review. Coordinate project scheduling
Shirajul Islam Shakur	Oversee data collection and preprocessing. Implementation.
Md. Mohaiminul Islam	Compile documentation and coordinate project experiments and testing.

11. Conclusion

In conclusion, General Graph Optimization (G2O) represents a significant advancement in the field of human pose estimation, particularly in localizing keypoints with enhanced precision and robustness. This thesis demonstrates that G2O's capability to effectively model and solve optimization problems over graph structures provides substantial benefits over traditional methods. The integration of G2O into keypoint localization frameworks addresses several critical challenges. It enhances the accuracy of keypoint detection by optimizing over graph-based representations, ensuring that the estimated keypoints adhere to both spatial and temporal constraints. This results in a more accurate and coherent depiction of human poses, even in the presence of occlusions, noise, and missing data. Moreover, G2O's ability to handle multi-person scenarios and adapt to various input modalities expands its applicability across different environments and conditions. Its efficiency in reducing computational complexity makes it suitable for real-time applications, facilitating its use in dynamic and resource-constrained settings.

12. References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] ZheCao, Gines Hidalgo, Tomas Simon, Shih-EnWei and YaserSheikh, ”OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, IEEE Transaction on Pattern Analysis and Machine Intelligence,arXiv:1812.08008v2 [cs.CV] 30 May 2019

[6] Anran Zhao, Jingli Li, Hongtao Zeng, Hongren Cheng, and Liangshan Dong,” DSPose: Dual-Space-Driven Keypoint Topology Modeling for HumanPoseEstimation ”, Received: 27 July 2023 Revised: 27 August 2023 Accepted: 1 September 2023 Published: 3 September 2023