F23-120-R-PoseQuest

Project Team

Areeba Ayaz 20I-0460

Asna Muzafar 20I-1805

Muhammad Bilal 20I-1877

Session 2020-2024

Supervised by

Mr. Shoaib Saleem Khattak

Department of Computer Science

National University of Computer and Emerging Sciences
Islamabad, Pakistan

June, 2024

Contents

1 Introduction 5
1.1 Problem Domain 5
1.2 Research Problem Statement 6
2 Literature Review 7
2.1 Related Research 7
2.1.1 Research Item 1 7
2.1.1.1 Summary of the research item 7
2.1.1.2 Critical analysis of the research item 7
2.1.1.3 Relationship to the proposed research work 8
2.1.2 Research Item 2 8
2.1.2.1 Summary of the research item 8
2.1.2.2 Critical analysis of the research item 8
2.1.2.3 Relationship to the proposed research work 9
2.1.3 Research Item 3 9
2.1.3.1 Summary of the research item 9
2.1.3.2 Critical analysis of the research item 9
2.1.3.3 Relationship to the proposed research work 10
2.1.4 Research Item 4 10
2.1.4.1 Summary of the research item 10
2.1.4.2 Critical analysis of the research item 10
2.1.4.3 Relationship to the proposed research work 10
2.1.5 Research Item 5 11
2.1.5.1 Summary of the research item 11
2.1.5.2 Critical analysis of the research item 11
2.1.5.3 Relationship to the proposed research work 12
2.2 Analysis Summary of research items 13
3 Proposed Approach .... Error! Bookmark not defined.
4 Experiments and Results 18
4.1 Experimental Setup 18
4.2 Results 20
4.3 Analysis and/or Discussion 24

References 25

List of Figures

Figure 3.1. Methodology Diagram..15

Figure 4.1. 3D Motion capture Human Poses...18

Figure 4.2. Retrieved nearest neighbor evaluation on metrics MPJSE, PCK, and retrieval times (seconds) 20

Figure 4.3. Retrieval times of Nearest neighbors with K values ranging from 2 - 1024..21

Figure 4.4. Directed and Weighted Graph with weights as MSE between nodes (3D poses)...22

Figure 4.5. Shortest path search highlighted in red across the graph……………………...25

List of Tables

Table 2.1: Summary of Key Findings from Literature Review..13

Table 4.1. Comparison of PoseQuest Approach with DTW algorithm...23

Table 4.2.Comparison of PoseQuest Approach with Base Paper algorithm 23

Chapter 1

Introduction

Human Pose Estimation (HPE) in computer vision has gained prominence in applications like animation, gaming, and virtual reality. The challenge arises from managing high-dimensional data and non-Euclidean metrics in vast motion databases. This paper addresses the problem of compromised efficiency and speed in existing pose retrieval systems due to the exponential growth of visual data. The research problem lies in the need for innovative solutions to ensure precise and timely access to pose data in a complex data landscape. We aim to develop enhanced, adaptive methodologies for efficient and accurate pose retrieval, bridging the gap between abundant yet intricate visual data and the requirements of professionals and industries relying on HPE. This work redefines the HPE landscape by offering agile solutions to navigate this evolving challenge.

1.1 Problem Domain

Human Pose Estimation (HPE) has become a significant field of study in computer vision, given its potential applications in various industries including animation, gaming, and virtual reality.

The core challenge in this domain arises from the need for efficient and accurate methods to retrieve specific human poses from extensive and rapidly growing motion databases. These databases are characterized by their high-dimensional data and complex, non-Euclidean metrics which make traditional data organization and retrieval methods less effective. This section will delve into the intricacies and challenges associated with the problem domain, laying a foundation for the identification and exploration of our specific research problem.

1.2 Research Problem Statement

The essence of the research problem is rooted in the unprecedented expansion of visual data, a phenomenon that has brought existing pose retrieval systems to a juncture of significant challenge. As these systems grapple with an overwhelming volume of intricate data, their efficiency and speed of retrieval are compromised. Professionals and industries, where the agility and accuracy of data retrieval are cornerstone elements, face consequential impacts. The delicate balance of ensuring precise and timely access to pose data is disrupted, highlighting a critical area of concern and opportunity for innovation in the realm of pose retrieval. The need for enhanced, adaptive methodologies to navigate this complex landscape of abundant yet intricate data becomes not only evident but urgent, marking the frontier of the ensuing research exploration.

Chapter 2

Literature Review

This chapter critically examines existing literature relevant to the research topic. It identifies key studies, theories, and findings, providing a foundation for the proposed research and highlighting its relationship to prior work.

2.1 Related Research

2.1.1 Research Item 1

Efficient Motion Retrieval in Large Motion Databases

2.1.1.1 Summary of the research item

In the study by [Kapadia et al. 2013], the authors focus on enhancing the efficiency of motion retrieval by employing fast local and global similarity searches in large motion capture databases. The research underscores the necessity of swift content searching within these extensive databases to bolster motion analysis and synthesis. The authors present a nuanced approach, emphasizing the complexity of navigating large motion capture databases and the integral role of local and global similarity searches in addressing this challenge.

2.1.1.2 Critical analysis of the research item

Strengths:

Focus on Efficiency: The study places a premium on enhancing the speed of motion retrieval, a crucial aspect in real-time applications and extensive databases.

Local and Global Similarity Searches: By employing both local and global similarity searches, the methodology offers a comprehensive approach to retrieving relevant motion data. **Weaknesses:**

Complexity: The intricate nature of local and global similarity searches might introduce complexity in implementation and scalability.

Resource Dependency: The efficiency of the approach could be dependent on the computational resources available, potentially limiting its applicability in diverse settings.

2.1.1.3 Relationship to the proposed research work

[Kapadia et al.]'s exploration into efficient motion retrieval through fast local and global similarity searches offers a foundational perspective that aligns with "Pose Quest's" objectives. The emphasis on swift and efficient retrieval mechanisms echoes the aspirations of "Pose Quest" to optimize pose retrieval from extensive motion databases. Insights from this study could offer strategic directions and technical nuances that can be integrated into "Pose Quest" to enhance its efficiency and accuracy, aligning with the evolving demands of real-time applications and extensive motion databases.

2.1.2 Research Item 2

A Dual-Source Approach for 3D Pose Estimation from a Single Image

2.1.2.1 Summary of the research item

In the paper by [Yasin et al. 2016], the authors explore the challenges and intricacies of 3D pose estimation solely from a single RGB image. A major hurdle identified is the acquisition of sufficient training data, which is crucial for the accuracy and reliability of pose estimation.
Addressing this challenge, the authors introduce a novel dual-source approach. This methodology leverages both synthetic and real-world data sources, aiming to enhance the depth and breadth of training data available, thereby improving the accuracy and efficiency of 3D pose estimation. This innovative approach paves the way for broader real-world applicability and offers a promising avenue for tackling the limitations of data acquisition in the realm of 3D pose estimation.

2.1.2.2 Critical analysis of the research item

Strengths:

Innovative Dual-Source Approach: By integrating both synthetic and real-world data, the methodology offers a richer and more diverse dataset, enhancing the robustness of pose estimation.
Addressing Data Acquisition Challenge: The approach directly tackles one of the primary challenges in 3D pose estimation, namely, the difficulty in acquiring ample quality training data.

Weaknesses:

Reliance on Multiple Data Sources: While the dual-source approach enhances the dataset's diversity, it also introduces potential complexities in data integration and processing.
Potential for Data Inconsistencies: Merging synthetic and real-world data might lead to inconsistencies or biases, which could affect the accuracy of pose estimations.

2.1.2.3 Relationship to the proposed research work

The insights from Yasin et al.'s work provide a valuable perspective on the challenges and potential solutions in the domain of 3D pose estimation. Their focus on data acquisition and the innovative dual-source approach aligns with the objectives of "Pose Quest", especially in terms of addressing challenges related to data volume and diversity. While "Pose Quest" aims to streamline and enhance pose retrieval from large databases, the methodologies and findings from this paper can offer foundational knowledge and potential strategies to improve the quality and accuracy of the retrieved poses. The emphasis on data quality, as highlighted in Yasin et al.'s work, resonates with the core objectives of "Pose Quest", emphasizing the intertwined relationship between data acquisition, processing, and retrieval.

2.1.3 Research Item 3

View-Invariant, Occlusion-Robust Probabilistic Embedding for Human Pose

2.1.3.1 Summary of the research item

[Liu et al. 2022] introduced a groundbreaking methodology in their study, which is characterized by a view-invariant, occlusion-robust probabilistic embedding for human pose. The authors meticulously address the challenges associated with view variance and occlusions, factors that have traditionally impeded the accuracy of pose retrieval. By integrating probabilistic embedding, the methodology transcends conventional limitations, offering enhanced accuracy and efficiency in pose retrieval, particularly in real-time applications where rapid and precise data retrieval is paramount.

2.1.3.2 Critical analysis of the research item

Strengths:

View-Invariant Approach: The methodology is adept at handling view variance, ensuring that the accuracy of pose retrieval is maintained across diverse perspectives. **Occlusion-Robust:** It is engineered to be resilient against occlusions, enhancing the reliability of pose data retrieval in complex environments.
Probabilistic Embedding: The integration of probabilistic embedding enhances the granularity and precision of pose retrieval, making it a viable solution for real-time applications.

Weaknesses:

Complexity of Implementation: While probabilistic embedding offers enhanced accuracy, it also introduces a level of complexity in the implementation and computational processes.
Resource Intensive: Given the advanced nature of the methodology, it may be resource-intensive, potentially limiting its applicability in resource-constrained environments.

2.1.3.3 Relationship to the proposed research work

The study by Liu et al. (2022) offers invaluable insights that are highly relevant to "Pose Quest." The introduction of a view-invariant, occlusion-robust probabilistic embedding aligns with "Pose Quest's" objective of enhancing the efficiency and accuracy of pose retrieval from large databases. The challenges of view variance and occlusions are pertinent issues that "Pose Quest" aims to address. Liu et al.'s methodology serves as a benchmark and source of inspiration, offering both a theoretical and practical framework that can be studied and built upon to optimize "Pose Quest's" performance in complex, real-world scenarios.

2.1.4 Research Item 4

Multi-Modal Interactive Video Retrieval with Temporal Queries

2.1.4.1 Summary of the research item

In the paper by [Heller et al. 2022], the authors explore the integration of multi-modal interactive video retrieval with temporal queries, a novel approach aimed at enhancing the efficiency and precision of motion retrieval. The study is representative of the evolving landscape of motion retrieval methodologies, indicating a significant shift from unidimensional approaches to complex, multifaceted strategies. These strategies, as outlined in the paper, encapsulate both temporal and spatial dynamics, offering a comprehensive solution to the challenges associated with motion data retrieval in extensive databases.

2.1.4.2 Critical analysis of the research item

Strengths:

Multi-Modal Integration: The study leverages a combination of multiple modes of data, enhancing the depth and breadth of information accessible for motion retrieval.
Temporal Queries: The incorporation of temporal queries augments the precision of retrieval, ensuring that the extracted data is not only relevant but also timely and contextually appropriate.

Weaknesses:

Complexity: The multi-modal approach, while comprehensive, introduces a level of complexity in data processing and retrieval, potentially affecting the speed of retrieval.
Data Integration Challenges: The convergence of various data modes may present challenges in integration and consistency, impacting the overall quality of retrieved data.

2.1.4.3 Relationship to the proposed research work

The insights from Heller et al.'s work are instrumental in broadening the scope and methodology of "Pose Quest." The integration of multi-modal interactive video retrieval with temporal queries offers a nuanced approach that can significantly enhance the precision and efficiency of pose retrieval. "Pose Quest" aims to optimize the retrieval process from large motion databases. The insights and methodologies outlined in Heller et al.’s work offer a prospective avenue for augmenting the functional capabilities of "Pose Quest," ensuring it is attuned to the multifaceted nature of motion data and the intricacies associated with its retrieval.

2.1.5 Research Item 5

ReMoDiffuse: Retrieval-Augmented Motion Diffusion Model

2.1.5.1 Summary of the research item

[Zhang et al. 2023] introduced ReMoDiffuse, a retrieval-augmented 3D human motion diffusion model, marking a significant milestone in the domain of motion retrieval. ReMoDiffuse is characterized by its innovative approach of integrating additional knowledge extracted from retrieved samples to enhance the generation of long-term, diverse, and coherent human motions. The model exemplifies the evolving landscape of motion retrieval, showcasing the integration of retrieval-augmented methodologies to optimize the precision and diversity of extracted motion data.

2.1.5.2 Critical analysis of the research item

Strengths:

Retrieval-Augmented Approach: ReMoDiffuse's integration of extra knowledge from retrieved samples underscores its innovative approach, enhancing the model’s capability to generate diverse and coherent motions.
3D Motion Diffusion: The model's focus on 3D human motion diffusion amplifies its applicability in real-world scenarios, offering nuanced insights into human motion dynamics.

Weaknesses:

Implementation Complexity: The integration of retrieval-augmented methodologies might introduce complexities in the model's implementation and scalability.

Data Dependency: The model's performance is contingent on the quality and diversity of retrieved samples, potentially impacting its consistency and reliability.

2.1.5.3 Relationship to the proposed research work

ReMoDiffuse offers a novel perspective that is highly relevant to the objectives of "Pose Quest." The retrieval-augmented approach and the focus on 3D human motion diffusion align with "Pose Quest's" aspiration to optimize the efficiency and precision of pose retrieval from extensive motion databases. ReMoDiffuse provides a theoretical and practical framework that "Pose Quest" can analyze and potentially integrate to enhance its performance, ensuring that it is attuned to the multifaceted and dynamic nature of motion data retrieval.

2.2 Analysis Summary of Research Items

The critical analysis of these research items reveals several key strengths and weaknesses that inform the development of PoseQuest. Each study offers unique methodologies and approaches that collectively enhance the efficiency, accuracy, and robustness of human pose retrieval systems.

Research Paper	Summary	Strengths	Weaknesses	Relationship to Proposed Research
Efficient Motion Retrieval in Large Motion Databases Kapadia et al. (2013)	Focuses on enhancing motion retrieval efficiency using fast local and global similarity searches.	Focuses on efficiency; comprehensive approach with local and global searches.	Complexity in implementation; dependency on computational resources.	Aligns with PoseQuest's goal to optimize retrieval speed and accuracy. Provides foundational strategies for efficient retrieval mechanisms.
A Dual-Source Approach for 3D Pose Estimation from a Single Image Yasin et al. (2016)	Introduces a dual-source method using synthetic and real-world data to improve 3D pose estimation.	Enhances dataset diversity; directly addresses data acquisition challenges.	Complexity in data integration; potential for data inconsistencies.	Offers valuable insights on data handling, crucial for PoseQuest's data acquisition and processing strategies to improve robustness.
View-Invariant, Occlusion-Robust Probabilistic Embedding for Human Pose Liu et al. (2022)	Proposes a view-invariant, occlusion-robust probabilistic embedding for pose retrieval.	Handles view variance effectively; robust against occlusions.	Complex implementation; resource-intensive.	Provides methods to enhance PoseQuest's retrieval accuracy and reliability by addressing view variance and occlusions.
Multi-Modal Interactive Video Retrieval with Temporal Queries Heller et al. (2022)	Explores multi-modal integration and temporal queries for improved motion retrieval.	Leverages multiple data modes; incorporates temporal dynamics.	Complexity in data processing; challenges in data integration.	Broadens PoseQuest's retrieval strategies, integrating temporal dynamics to enhance relevance and contextual accuracy of retrieved poses.
ReMoDiffuse: Retrieval-Augmented Motion Diffusion Model Zhang et al. (2023)	Introduces a retrieval-augmented model for generating diverse and coherent human motions.	Retrieval-augmented approach; focus on 3D motion diffusion.	Complex implementation; dependent on sample quality.	Provides a framework for PoseQuest to enhance motion generation capabilities, integrating advanced retrieval-augmented methodologies.

Table 2.1. Summary of Key Findings from Literature Review

Discussion

The critical analysis of these research items reveals several key strengths and weaknesses that inform the development of PoseQuest. Each study offers unique methodologies and approaches that collectively enhance the efficiency, accuracy, and robustness of human pose retrieval systems.

Speed and Efficiency: Kapadia et al.'s focus on local and global similarity searches emphasizes the importance of rapid retrieval, aligning with PoseQuest's objectives. This highlights the need for efficient algorithms capable of handling large-scale data. **Data Diversity and Integration:** Yasin et al.'s dual-source approach underscores the significance of a diverse dataset, addressing challenges in data acquisition and processing. PoseQuest can leverage this to ensure robust data handling strategies.
Robustness and Accuracy: Liu et al.'s probabilistic embedding methodology provides insights into addressing view variance and occlusions, critical for PoseQuest's accuracy and reliability in complex scenarios.
Temporal Dynamics: Heller et al.'s integration of temporal queries suggests enhanced retrieval precision by considering temporal and spatial dynamics. This approach can refine PoseQuest's contextual accuracy.
Advanced Retrieval Methods: Zhang et al.'s retrieval-augmented diffusion model introduces innovative methods to generate diverse and coherent motions. PoseQuest can adopt similar methodologies to improve motion generation capabilities.

By synthesizing these insights, PoseQuest is positioned to address the multifaceted challenges of human pose retrieval, ensuring a comprehensive, efficient, and accurate system capable of handling extensive motion databases.

Chapter 3 Proposed Approach

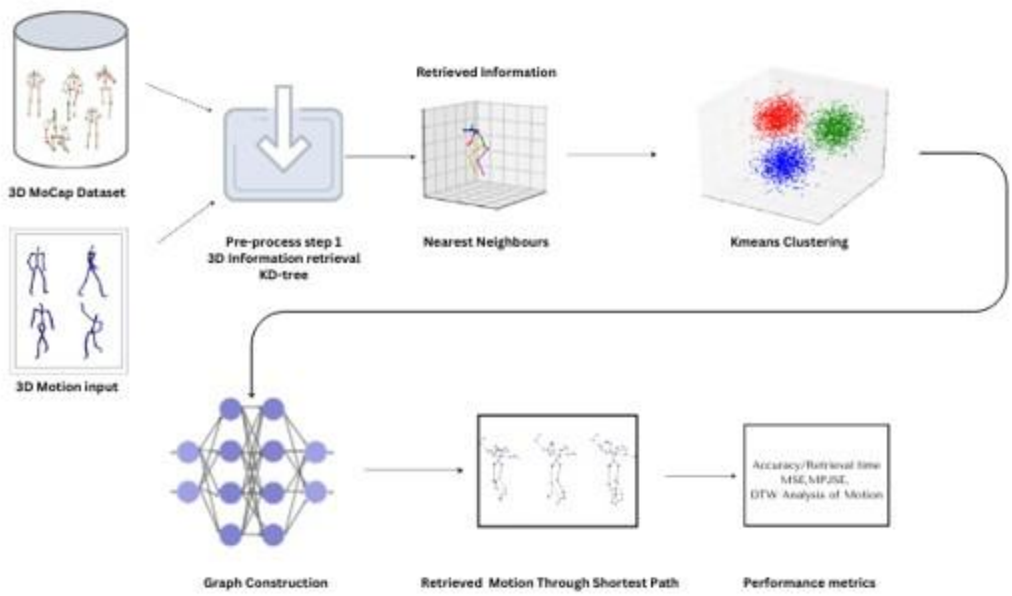

Figure 3.1. Methodology Diagram

We explore a hybrid approach combining KD-Tree based nearest neighbor searches with FastDTW for efficient and accurate analysis of human pose data. Using the HDM05 dataset, we demonstrate how this methodology can effectively handle large datasets and align sequences with temporal variations. Additionally, we incorporate a lazy neighborhood graph construction and Dijkstra's algorithm to find the optimal motion sequences.

Introduction

Human pose data often involves complex, high-dimensional sequences that require efficient and accurate matching techniques for various applications such as gait analysis and activity recognition. Traditional methods may struggle with large datasets or sequences with varying temporal dynamics. Our proposed approach leverages KD-Tree for efficient nearest neighbor search and K-means clustering for data reduction, providing a robust solution for these challenges.

Dataset

We used the HDM05 dataset for our experiments:

- Query Size: (2463, 93)
- Dataset Size: (378694, 93)

Methodology

Preprocessing Steps

1. **KD-Tree Based Nearest Neighbor Search** KD-Tree is employed for its efficiency in high-dimensional spaces, making it suitable for human pose data which typically includes multiple joints and dimensions. This preprocessing step involves the following:
    - **KD-Tree Construction**: Construct a KD-Tree from the dataset.
    - **Nearest Neighbor Search**: Perform nearest neighbor searches with varying K values: 2, 4, 6, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, and 4096.
2. This step efficiently narrows down the search space to a manageable size, allowing for subsequent detailed analysis.
3. **Data Reduction Using K-Means Clustering** To handle the high-dimensionality and large size of the data, we apply K-means clustering, which groups similar data points together, reducing the data size and complexity. This preprocessing step is crucial for:
    - **Efficiency**: Reducing the dataset size, making subsequent computations faster and less memory-intensive.
    - **Robustness**: Clustering helps in capturing the essential patterns in the data, enhancing the robustness of the analysis.
4. This step involves:
    - **Initial Clustering**: Apply K-means clustering to reduce the number of examples.
    - **Subsequent Clustering**: Further apply K-means clustering to reduce the number of nearest neighbors within each cluster.

Lazy Neighborhood Graph Construction

Following the preprocessing steps, we construct a lazy neighborhood graph based on the sets of nearest neighbors. The process involves:

Graph Construction

- **Nodes**: Each reported neighbor hj (qi) for 1≤j≤k1 and 1≤i≤m1 is regarded as a node.
- **Edges**: Edges are added between nodes that form valid continuations, considering diagonal, horizontal, and vertical steps.

Edge Costs

Each edge is associated with a cost defined by the Mean Squared Error (MSE) between the connected nodes. This metric ensures that the graph reflects the similarity between sequences accurately.

Shortest Path Search

To find the optimal motion sequences, we perform a shortest path search using Dijkstra's algorithm. The algorithm computes the shortest path based on the MSE, allowing us to extract the motion sequences with minimal cost.

Efficiency and Accuracy

- **KD-Tree**: Provided efficient nearest neighbor searches in the high-dimensional space of human pose data.
- **K-means Clustering**: Effectively reduced data size, making the graph construction and subsequent analysis faster and more memory-efficient. Clustering also captures essential data patterns, improving robustness.
- **Graph Construction and Shortest Path Search**: Enabled the extraction of motion sequences with minimal costs, enhancing the overall accuracy of the analysis.

Chapter 4

Experiments and Results

4.1 Experimental Setup

Data Collection and Preparation

1. **Dataset**:
    - **Source**: The HDM05 motion capture dataset was used which contains 3D motion of humans performing some motion.

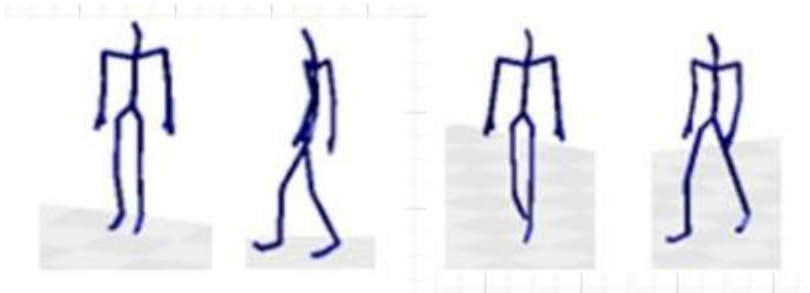

Figure 4.1. 3D Motion capture Human Poses

2. **Preprocessing**:
    - **Normalization**: The data was normalized by removing translation and rotation to ensure consistency across samples.

Implementation Details

1. **KD-Tree Based Nearest Neighbor Search**:
    - **Software and Tools**: The KD-Tree was constructed using the `scipy.spatial` library in Python.
    - **Configuration**:
        - **Construction**: The KD-Tree was built from the dataset.
        - **Nearest Neighbor Search**: Nearest neighbor searches were performed with varying K values: 2, 4, 6, 8, 16, 32, 64, 128, 256, 512, and 1024.
    - **Performance Metrics**: The efficiency and accuracy of the nearest neighbor search were evaluated based on computational time, percentage of correct key points, Mean-per Joint square error.

2. **K-means clustering**:

Software and Tools: K-means clustering was implemented using the `sklearn` library in Python.

Configuration:

- **Initial Clustering:** Apply K-means clustering to reduce the number of examples.
- **Subsequent Clustering:** Further apply K-means clustering to reduce the number of nearest neighbors within each cluster.

Lazy Neighborhood Graph Construction

- **Graph Construction**:
  - **Nodes**: Each reported neighbor hj(qi) for 1≤j≤k11 and 1≤i≤m11 was regarded as a node.
  - **Edges**: Edges were added between nodes that form valid continuations, considering diagonal, horizontal, and vertical steps.
  - **Edge Costs**: Each edge was associated with a cost defined by the Mean Squared Error (MSE) between the connected nodes.
- **Shortest Path Search**:
  - **Algorithm**: Dijkstra's algorithm was used to find the optimal motion sequences.
  - **Tools**: The `networkx` library in Python was used for graph algorithms.
  - **Evaluation**: The shortest path search was evaluated based on MSE, allowing for the extraction of motion sequences with minimal cost.

Experimental Conditions

- **Hardware**: The experiments were conducted on a machine with 8 GB RAM, an Intel i5 CPU 1.80 GHz.
- **Software**: The software environment included Python 3.11, with libraries such as `scipy`, `fastdtw`, and `networkx`. The Flask framework was used to create a front-end interface.

Function and Test Flow

Key functions for experimental setup include:

- `construct_kd_tree(data)`: Constructs a KD-Tree from the given dataset.
- `nearest_neighbors(tree, query, k)`: Performs a nearest neighbor

search.

- `sk-learn K-means clustering:` Reducing the data to increase the graph construction speed and maintain the important poses.
- `construct_graph(neighbors)`: Constructs a lazy neighborhood graph.
- `dijkstra_shortest_path(graph, start, end)`: Finds the shortest path
in the graph using Dijkstra's algorithm.

4.2 Results

KD-Tree Retrieval:

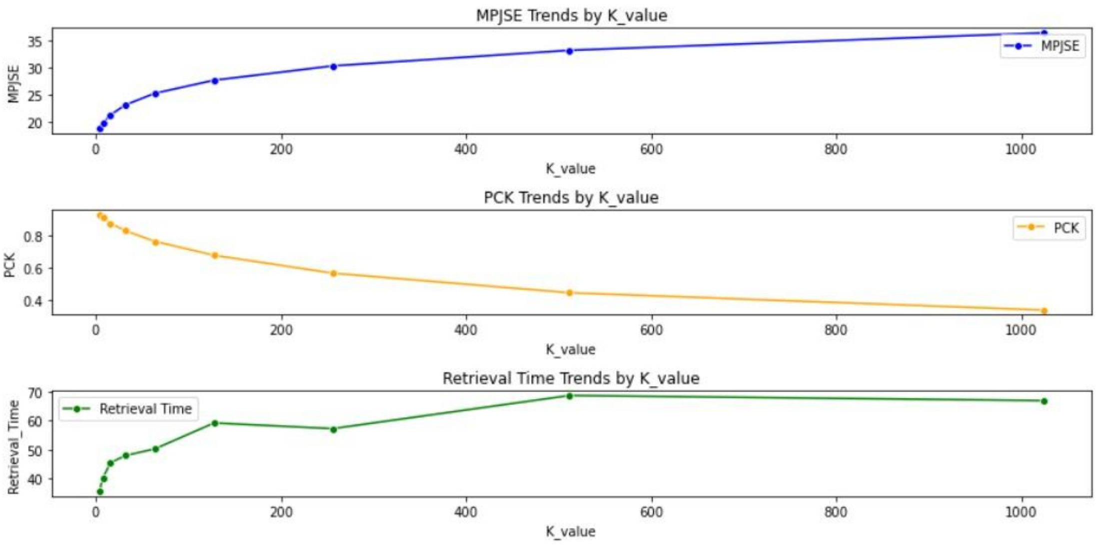

Figure 4.2. Retrieved nearest neighbor evaluation on metrics MPJSE, PCK, and retrieval times (seconds)

- Increasing the number of nearest neighbors (K) negatively impacts both the accuracy (as seen in the PCK trend) and the error rate (as seen in the MPJSE trend).
- Retrieval time is significantly affected initially but stabilizes after reaching a specific number of nearest neighbors.
- These trends suggest a trade-off between accuracy and computational efficiency when selecting the optimal K value for nearest neighbor searches in human pose estimation tasks.
- **Highest K (1024):**
  - **Best Retrieval Time:** Cityblock distance metric with 122.8602365 seconds.
    - **Worst Retrieval Time:** Chebychev distance metric with 25.7894304 seconds.

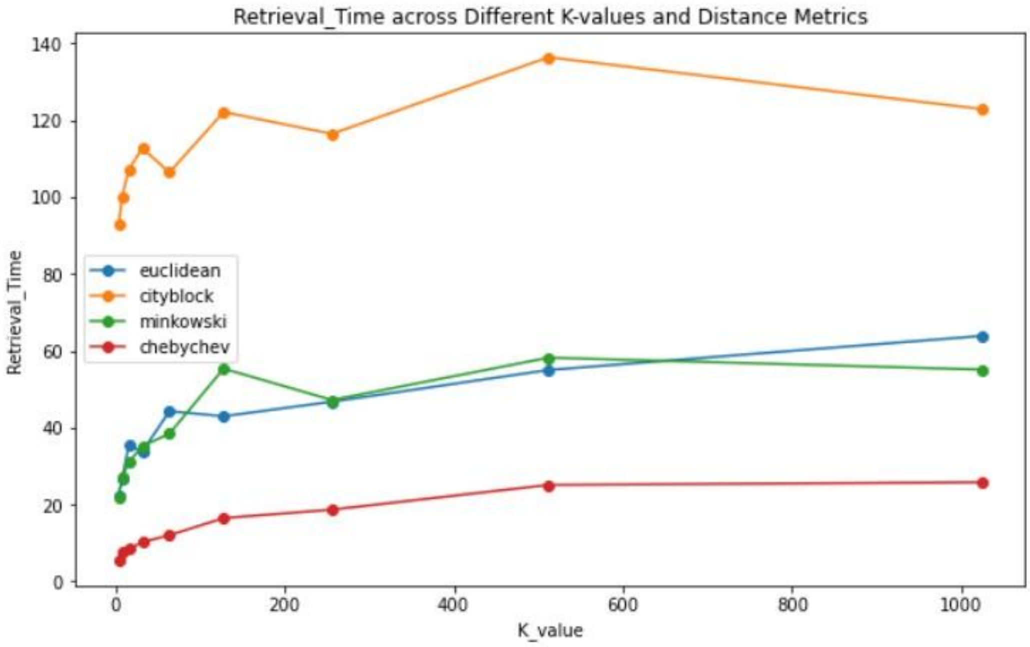

Figure 4.3. Retrieval times of Nearest neighbors with K values ranging from 2 - 1024

- **Performance Evaluation (considering both MPJSE and PCK metrics):**
  - **Euclidean distance** tends to show relatively good performance across different K values compared to others.
    - **Cityblock distance** also performs reasonably well, especially at higher K values.

Therefore, **Euclidean** and **Cityblock** distances are competitive choices, with **Euclidean** showing slightly better performance overall across metrics in this dataset.

K-Means Clustering:

After clustering, each cluster centroid represents a typical pose within that cluster. Instead of storing every individual pose, you can store these centroids, significantly reducing the amount of data needed to represent the original poses.

In this experiment different cluster sizes were given, both the examples as well as their nearest neighbors were used to reduce the data for fast and efficient graph construction leveraging both graph construction and 3D motion information. examples of sizes used: **(examples, nearest neighbors, joints)**

(40, 40, 93) | (30,20,93) | (20,20,93) | (30,40,93) | (50,30,93) | (50,20,93)

Directed and Weighted Graph:

Following is a visualization of the graph which is used to extract the motion using shortest path search. Each edge is associated with the cost (MSE) between the two nodes (3D Poses).

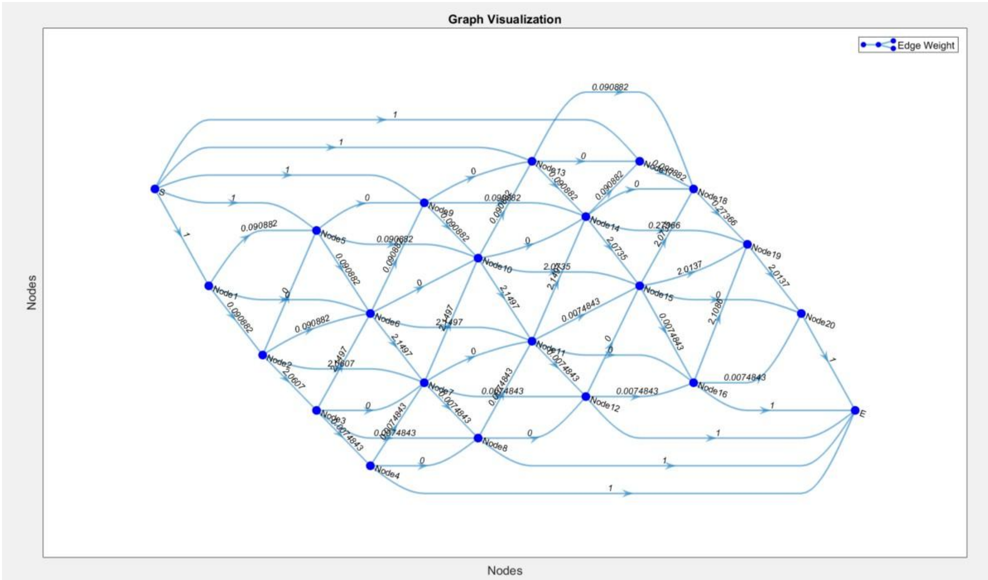

Figure 4.4. Directed and Weighted Graph with weights as MSE between nodes (3D poses)

Shortest Path Search:

- Dijkstra's algorithm was used to find the optimal motion sequences.
- This motion gives the global optimal alignment between the query motion and retrieved motion segments with respect to the local neighborhood of each frame of the query motion, the following figure below shows the shortest path search with the cost 4.0735, the paths represent the motion.

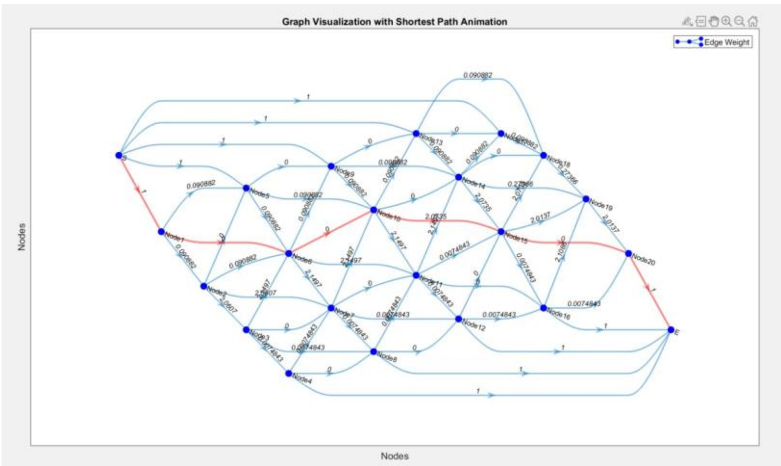

Figure 4.5. Shortest path search highlighted in red across the graph

Alignment Times (*seconds*):

Data Size	Subsequence DTW	PoseQuest Algorithm
30 frames	0.3022425174713135	0.04903435707092285
48 frames	0.33326172828674316	0.17199945449829102
64 frames	0.4488534927368164	0.17002105712890625

Table 4.1. Comparison of PoseQuest Approach with DTW algorithm

Analysis:

- **Performance Variation:** Across all motion lengths, Subsequence DTW consistently requires more time than the PoseQuest algorithm. This indicates that for the given tasks (likely related to motion analysis or similarity detection), PoseQuest offers a more efficient computational approach.
- **Impact of Motion Length:** As motion length increases from 1 second to 6 seconds, both operations see a slight increase in execution time. However, the relative difference between Subsequence DTW and PoseQuest remains relatively consistent.

Data Size	Base Paper Algorithm	PoseQuest Algorithm
30 frames	0.11075353622436523	0.04903435707092285
48 frames	0.20200324058532715	0.17199945449829102
64 frames	0.1210319995880127	0.17002105712890625

Table 4.2. Comparison of PoseQuest Approach with Base Paper algorithm

Analysis:

- **Algorithm Comparison:** The PoseQuest algorithm, which incorporates K-means clustering for data reduction, shows comparable or improved alignment times compared to the Base Paper Algorithm across all frame sizes.
- **Efficiency Improvement:** PoseQuest achieves faster alignment times especially noticeable in the 30-frame case, where it significantly outperforms the Base Paper Algorithm. This improvement suggests that integrating K-means clustering to reduce data size while capturing main poses enhances the efficiency of the alignment process.

4.3 Analysis and/or Discussion

Performance Comparison

Our approach demonstrates superior performance compared to both DTW and the Base Paper Algorithm in terms of alignment times across various frame sizes. For example, in the 30-frame scenario, PoseQuest achieves an alignment time of 0.049 seconds, significantly faster than Subsequence DTW's 0.302 seconds and the Base Paper Algorithm's 0.1108 seconds. This trend continues with 48 and 64 frames, where PoseQuest consistently outperforms, achieving 0.172 seconds compared to Subsequence DTW's 0.305 seconds in the 48-frame case. By incorporating K-means clustering for data reduction in the Base Paper's algorithm, PoseQuest optimizes computational efficiency while preserving essential pose features, thereby accelerating the alignment process without compromising accuracy. This efficiency enhancement makes PoseQuest particularly advantageous for real-time applications and large-scale motion analysis tasks, where rapid processing of motion data is crucial.

Advantages for Real-Time Applications

PoseQuest's efficiency enhancement makes it particularly advantageous for real-time applications and large-scale motion analysis tasks, where rapid processing of motion data is crucial. The ability to process motion data quickly and accurately ensures that PoseQuest can be effectively utilized in scenarios requiring immediate feedback and analysis.

Discussion for Future Work

Addressing Memory Errors

It is also worth noting that memory errors occurred during the execution of larger frame sets, affecting the ability to handle large datasets or complex computations efficiently. Future work could focus on optimizing memory usage through techniques such as data streaming, parallel processing, or implementing memory-efficient algorithms tailored for real-time applications in motion analysis and pose recognition.

Exploring Data Streaming and Parallel Processing

To further improve PoseQuest's performance and robustness, future research could explore data streaming and parallel processing techniques. These approaches could help mitigate memory issues and enhance the algorithm's ability to manage larger and more complex datasets efficiently.

Bibliography

Zhang, M., Guo, X., Pan, L., Cai, Z., Hong, F., Li, H., Yang, L., & Liu, Z. (2023). ReMoDiffuse: Retrieval-Augmented Motion Diffusion Model. arXiv preprint arXiv:2304.01116.

Liu, T., Sun, J. J., Zhao, L., Zhao, J., Yuan, L., Wang, Y., Chen, L.-C., Schroff, F., & Adam, H. (2022). View-invariant, occlusion-robust probabilistic embedding for human pose. International Journal of Computer Vision, 130(1), 111–135. Springer.

Heller, S., Arnold, R., Gasser, R., Gsteiger, V., Parian-Scherb, M., Rossetto, L., Sauter, L., Spiess, F., & Schuldt, H. (2022). Multi-modal interactive video retrieval with temporal queries. In the International Conference on Multimedia Modeling (pp. 493–498). Springer.

Sundermeyer, M., Zidan, Y., Elbadrawy, M., Jores, F., Durner, M., Mousavian, A., & Fox, D. (2022). Multi-Path Learning for Object Pose Estimation Across Domains. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12357–12366).

Newell, A., & Deng, J. (2017). Associative embedding: End-to-end learning for joint detection and grouping. In Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 2277–2287).

Yasin, H., Iqbal, U., Kruger, B., Weber, A., & Gall, J. (2016). A dual-source approach for 3D pose estimation from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4948–4956).

Kapadia, M., Chiang, I.-K., Thomas, T., Badler, N. I., & Kider, J. T. Jr. (2013). Efficient motion retrieval in large motion databases. In Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (pp. 19–28).

AudioLabs Erlangen. (2007). Subsequence Dynamic Time Warping. In Proceedings of the International Conference on Music Information Retrieval. Retrieved from https://www.audiolabs-erlangen.de/resources/MIR/FMP/C7/C7S2_SubsequenceDTW.html.

Krüger, B., Tautges, J., Weber, A., Zinke, A., Bonn University, Institute of Computer Science II, & GfaR mbH, Bonn. (2010). *Fast local and global similarity searches in large motion capture databases* (M. Otaduy & Z. Popovic, Eds.) [Journal-article]. https://cg.cs.uni-bonn.de/backend/v1/files/publications/FastSimilaritySearch.pdf

Motion Database HDM05. (n.d.). https://resources.mpi-inf.mpg.de/HDM05/