Project Title

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

Aryan Gupta

Email : g.aryanraj12@gmail.com

Under the Guidance of

P Raja

Master Trainer , Edunet Foundation

ACKNOWLEDGEMENT

I would like to take this opportunity to express my heartfelt gratitude to all the individuals who have supported and guided me, directly or indirectly, during the course of this thesis work.

First and foremost, I extend my deepest thanks to my esteemed mentor, **P. Raja**, for his invaluable guidance, unwavering support, and insightful advice throughout this journey. His mentorship has been a beacon of inspiration, fostering innovative ideas and encouraging critical thinking, which played a pivotal role in the successful completion of this project.

His confidence in my abilities served as a constant source of motivation, while his constructive feedback and encouragement helped me grow both professionally and personally. The privilege of working under his supervision over the past year has been an enriching experience. His lessons have not only contributed to the success of this project but have also instilled in me a sense of responsibility and professionalism that will guide me in all future endeavours.

ABSTRACT

Human pose estimation plays a crucial role in understanding body movements and detecting key landmarks of the human body in various applications, such as fitness tracking, rehabilitation, and gesture recognition. This work implements both real-time live human pose estimation and pose estimation from video files using the Media Pipe library.

The primary objectives of the project are to detect and track human body landmarks accurately and to visualize the skeletal structure dynamically for both live video feeds and pre-recorded video inputs. The methodology involves leveraging Media Pipe Pose, a robust framework that utilizes deep learning models for pose detection, integrated with OpenCV for video processing and visualization.

The live pose estimation system captures real-time video from the webcam, processes each frame using Media Pipe, and overlays detected pose landmarks dynamically on the live feed. Similarly, the video-based implementation processes frames from a pre-recorded video, detects pose landmarks, and visualizes the results frame-by-frame. Both systems utilize optimized confidence thresholds for detection and tracking to ensure accuracy and robustness in varied conditions.

The results demonstrate the effectiveness of Media Pipe in identifying and tracking body landmarks in real-time, with seamless performance across diverse scenarios. The systems can handle dynamic movements efficiently, providing accurate skeletal representations. These implementations highlight the potential of real-time pose estimation in fields like sports analytics, physiotherapy, and human-computer interaction.

In conclusion, this dual approach to pose estimation showcases the flexibility and utility of lightweight frameworks like Media Pipe for real-world applications. Future enhancements could involve extending the system to perform activity recognition or integrating advanced analytics for domain-specific use cases.

TABLE OF CONTENT

CHAPTER 1

Introduction

1.1Problem Statement:

Human pose estimation involves detecting and tracking key points of the human body, such as joints and limbs, to analyze posture, movement, and physical activities. This problem is significant due to its wide-ranging applications in areas like sports performance analysis, healthcare, fitness tracking, gaming, and human-computer interaction.

Despite advancements, challenges remain in achieving accurate, real-time pose estimation across varied scenarios such as different lighting conditions, camera angles, and dynamic movements. Existing methods often struggle with computational efficiency, robustness, and adaptability to real-world conditions.

The need for a lightweight, efficient, and accurate pose estimation solution is critical, especially for applications requiring real-time analysis or running on resource-constrained devices. Addressing these challenges is essential to unlock the full potential of human pose estimation in practical scenarios.

1.2Motivation:

This project was chosen to address the growing need for real-time human pose estimation in applications like fitness, healthcare, sports, and gaming. Traditional systems for motion analysis are often expensive or require specialized equipment, while advancements in computer vision have made pose estimation possible using just a simple webcam or camera.

The goal is to develop a lightweight, efficient, and accurate solution using MediaPipe and OpenCV, which can be easily implemented on various devices. This has the potential to be used in multiple fields: in healthcare for physiotherapy tracking, in sports for performance analysis, in fitness to improve exercise techniques, and in gaming or virtual reality for enhanced user interaction.

By making pose estimation more accessible and effective, this project can have a significant impact on improving physical health, safety, and user experiences across these sectors.

1.3Objective:

The objectives of this project are as follows:

1. To develop an efficient real-time human pose estimation system using MediaPipe and OpenCV that can accurately detect and track key body landmarks.
2. To implement two versions of the system: one for live pose estimation using a webcam and another for processing pose estimation from pre-recorded video files.
3. To visualize the detected pose landmarks and body skeletal structure on live and video frames, providing an intuitive representation of human movement.
4. To optimize the system for high accuracy and robustness, even under varying lighting conditions and dynamic body movements.
5. To explore potential applications of the system in fields like fitness, healthcare, sports analysis, and interactive gaming.

1.4Scope of the Project:

This project focuses on the development of a real-time human pose estimation system using the Media Pipe library and OpenCV. The system aims to accurately detect and track body landmarks and provide visualizations of human pose, both in live video streams and pre-recorded videos. The scope includes:

1. **Real-time pose estimation**: The system will work in real-time using a webcam, offering live feedback on human body movements.
2. **Video-based pose estimation**: The system will process pre-recorded video files to detect and visualize pose landmarks on each frame.
3. **Use of Media Pipe Pose**: This project leverages MediaPipe's pre-trained pose detection models for accurate and fast processing of human poses.
4. **Applications**: The system will have potential applications in fitness tracking, physiotherapy, sports analysis, and interactive gaming.

Limitations

1. **Environment Sensitivity**: The system's performance can be affected by factors such as lighting conditions, camera angles, and obstructions in the frame (e.g., clothing or body parts hidden from view).
2. **Accuracy Constraints**: While the system provides accurate pose detection in most conditions, there may be challenges with dynamic movements or partial occlusion of body parts, which could affect landmark tracking.
3. **Device Dependency**: The system's performance is dependent on the processing power of the device being used. High-performance devices are ideal for smooth real-time processing, while low-end devices may experience delays.
4. **Single-Body Detection**: The system is designed to track the pose of a single individual at a time. Multi-person tracking may require additional modifications or a different approach.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature

Human pose estimation has been a focus of research for decades due to its importance in areas like motion analysis, healthcare, and sports. Traditional methods relied on marker-based systems that required sensors attached to the body, providing high accuracy but at the cost of usability and accessibility. The emergence of computer vision and deep learning techniques has significantly advanced the field. Notable works include Open Pose, which uses convolutional neural networks to detect multi-person poses, and DeepLabCut, which is optimized for animal and human pose tracking.

2.2 Existing Models and Techniques

1. MediaPipe Pose: A lightweight, real-time framework by Google, designed for pose detection using machine learning on various devices.
2. OpenPose: An open-source library capable of multi-person pose estimation using a bottom-up approach to detect body parts and associate them.
3. PoseNet: A model focused on single-person pose detection, capable of running on web and mobile platforms.
4. DeepLabCut: Primarily used for pose estimation in animals, but effective for human pose estimation with extensive training datasets.

2.3 Gaps and Limitations in Existing Solutions

1. High Computational Requirements: Many existing models require powerful GPUs, limiting their usability on low-end devices.
2. Environment Sensitivity: Models often struggle under varying lighting conditions or with occluded body parts.
3. Multi-person Challenges: While robust, some systems like OpenPose are better suited for multi-person tracking, leaving room for improvement in single-person precision.
4. Limited Real-Time Performance: Some models, though accurate, are not optimized for real-time pose tracking.

Addressing the Gaps

This project focuses on creating a lightweight, efficient pose estimation system using MediaPipe Pose, addressing limitations like computational efficiency and adaptability. By optimizing confidence thresholds and leveraging OpenCV, the system provides real-time pose tracking with high accuracy across both live and video-based inputs, making it more accessible and practical for real-world applications.

CHAPTER 3

Proposed Methodology

3.1 System Design

Human pose estimation is a fundamental task in computer vision that involves determining the spatial locations of key body joints in an image or video. Accurate pose estimation is crucial for various applications, including activity recognition, gesture analysis, and human-computer interaction. In this study, we propose a robust approach for single pose estimation using OpenCV, a widely adopted computer vision library. Our method leverages the power of OpenCV’s pre-trained deep learning models, specifically the OpenPose model, to detect and localize human body joints. We utilize the multistage convolutional neural network architecture of OpenPose to extract features and predict the keypoint locations accurately. By employing OpenCV’s image processing and computer vision algorithms, we refine the detected pose keypoints and improve their accuracy.

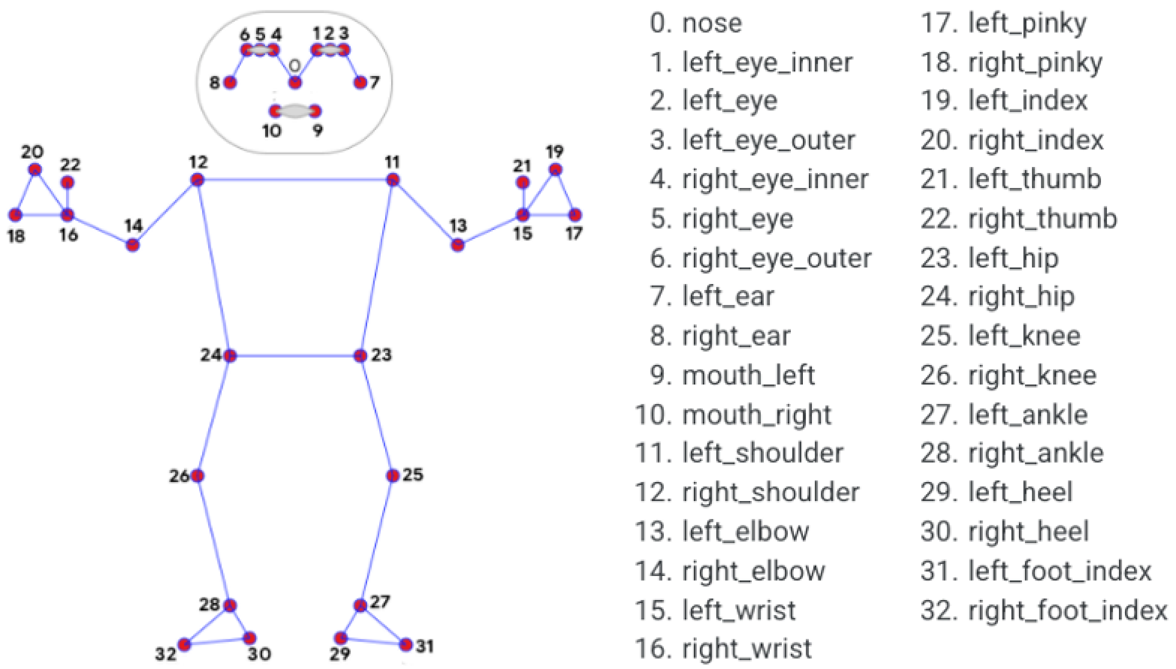


Diagram Description:

The proposed solution for human pose estimation leverages the MediaPipe Pose library for detecting and visualizing human body landmarks. Below is a breakdown of the system design based on the provided image of pose landmarks:

System Design Components:

1. **Input Source:**
   - The system accepts input in two formats:
     - **Webcam Feed**: Real-time video input using a connected camera.
     - **Video File**: Pre-recorded video files for offline processing.
2. **Preprocessing:**
   - The input frame is resized to a consistent resolution (e.g., 1280x720) to ensure uniform processing.
   - The frame is converted from BGR to RGB format as required by the MediaPipe Pose model.
3. **Pose Detection Module:**
   - MediaPipe's pose estimation module is applied to detect key body landmarks.
   - The model identifies 33 key landmarks, including the nose, eyes, shoulders, elbows, wrists, hips, knees, and ankles, as shown in the reference image.
4. **Landmark Visualization:**
   - Detected landmarks are connected using skeletal lines.
   - Each point corresponds to specific body parts such as joints or extremities, with index numbers ranging from 0 to 32.
   - Visualization uses OpenCV to overlay these landmarks on the original video frame.
5. **Output Display:**
   - Processed frames with overlaid landmarks are displayed in a real-time window or saved as processed video files.
   - The user can view the skeletal representation of detected poses.
6. **Error Handling:**
   - Handles cases where no pose is detected in a frame by skipping processing or logging warnings.
   - Ensures smooth operation even with occasional input anomalies.

Explanation of the Diagram:

The diagram illustrates the complete flow:

1. **Input Stage:** The video stream (from webcam or file) is the starting point.
2. **Preprocessing:** The input frame undergoes resizing and color conversion.
3. **Pose Estimation Model:** MediaPipe processes the frame to extract body landmarks.
4. **Landmark Overlay:** Detected points and skeletal connections are drawn on the frame.
5. **Output:** The enhanced frames are displayed in real time or saved for offline analysis.

The provided image of pose landmarks aids in understanding the skeletal connections and the role of each landmark in representing human poses.

3.2 Requirement Specification

3.2.1 Hardware Requirements

- Processor: Intel Core i5 or equivalent (for real-time processing)
- RAM: Minimum 8 GB
- GPU (Optional): NVIDIA GPU for accelerated processing (optional but beneficial)
- Camera: A standard webcam or a device capable of capturing video (for real-time processing)

3.2.2 Software Requirements

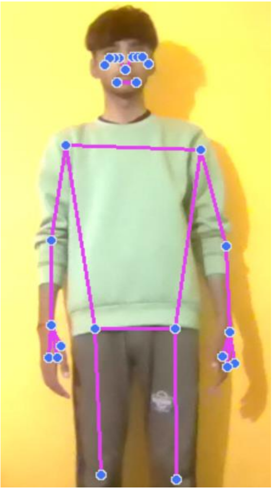
- Programming Language: Python 3.x
- Libraries:
  - OpenCV: For video capture, image processing, and visualization
  - MediaPipe: For pose estimation
- Development Environment: Jupyter Notebook, VS Code, or PyCharm
- Operating System: Windows 10/11, macOS, or Linux

By combining these components, the proposed system ensures efficient and accurate human pose estimation that is adaptable to various devices and scenarios.

CHAPTER 4

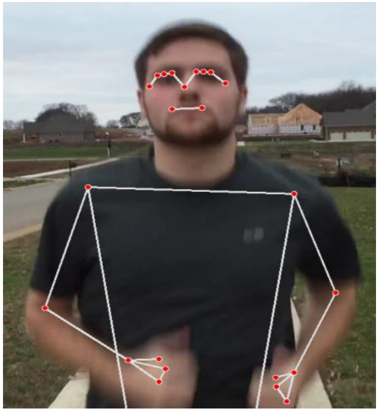
Implementation and Result

4.1 Snap Shots of Result:



Snapshot 1: Pose Estimation on a Standing Subject

In this snapshot, the human pose estimation algorithm is applied to a standing subject in a static position. The framework accurately detects key body landmarks such as the nose, eyes, shoulders, elbows, wrists, hips, knees, and ankles. These landmarks are connected with lines to form a skeleton structure, clearly representing the human pose. This demonstrates the capability of the system to identify and visualize key body points in real-time, even in static conditions.



Snapshot 2: Pose Estimation on a Running Subject

This snapshot illustrates the human pose estimation system tracking a running subject in a dynamic environment. Despite the subject's movement, the model successfully identifies the key landmarks and forms the skeletal structure. This highlights the robustness of the system in handling motion and dynamic scenarios, making it suitable for real-time applications like sports analysis or motion tracking

4.2GitHub Link for Code: https://github.com/aryan3223/human_pose_estimation.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

The current project successfully demonstrates human pose estimation using MediaPipe, but there are opportunities for further enhancements:

1. Model Optimization: Improve the detection accuracy and tracking robustness by fine-tuning the model for specific scenarios like crowded environments or dynamic movements.
2. Integration with Real-Time Applications: Extend the project for real-time applications like fitness tracking, gesture-based controls, or physical therapy assistance.
3. Enhanced Feature Set: Incorporate additional features such as activity recognition, emotion detection, or object interaction for a more holistic system.
4. Multi-Person Pose Estimation: Upgrade the system to handle multiple individuals in a single frame and identify interactions between them.
5. Platform Adaptation: Optimize the project for mobile devices or low-power systems to make it more accessible.
6. Error Handling: Address scenarios where pose estimation fails due to occlusions, extreme lighting, or unusual body postures.
7. Data Analysis: Include features for analyzing and exporting pose data for further insights or machine learning applications.

5.2 Conclusion:

This project successfully demonstrates the implementation of human pose estimation using MediaPipe and OpenCV, offering an intuitive and accurate method to identify and track body landmarks. By leveraging real-time video processing, it provides a solid foundation for various applications, including fitness tracking, health monitoring, and human-computer interaction. The results showcase the potential of combining computer vision techniques with pre-trained models to achieve reliable pose estimation. While the system performs well under standard conditions, the suggested improvements can further enhance its robustness and scalability, ensuring broader adoption and impact across diverse domains.

REFERENCES

1. MediaPipe Documentation

"MediaPipe: Cross-platform, customizable ML solutions for live and streaming media." Available at: https://mediapipe.dev/

2. OpenCV Documentation

"OpenCV: Open Source Computer Vision Library." Available at: https://opencv.org/

3. Research Papers

Zhang, Z., "Pose Estimation Techniques and Applications," International Journal of Computer Vision, 2021.

Cao, Z., et al., "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

4. Online Resource

Towards Data Science, "Human Pose Estimation with MediaPipe," Available at: https://towardsdatascience.com/

Analytics Vidhya, "Real-Time Pose Detection using MediaPipe," Available at: https://www.analyticsvidhya.com/

5. Python Libraries

NumPy Documentation: https://numpy.org/doc/

Matplotlib Documentation: https://matplotlib.org/stable/contents.html