# Advanced Upper Body Detection and Metric Analysis Using MediaPipe

*Prof. Ganesh Kadam*
*Dept. of Computer Engineering*
*Pimpri Chinchwad College Of Engineering*
*Pune, India*
*ganesh.kadam@pccoepune.org*

*Aryan Baheti*
*Computer Engineering*
*Pimpri Chinchwad College Of Engineering*
*Pune, India*
*aryan.baheti21@pccoepune.org*

*Akshay Chaudhari*
*Computer Engineering*
*Pimpri Chinchwad College Of Engineering*
*Pune, India*
*akshay.chaudhari21@pccoepune.org*

*Amogh Chandragiri*
*Computer Engineering*
*Pimpri Chinchwad College Of Engineering*
*Pune, India*
*amogh.chandragiri21@pccoepune.org*

***Abstract*— Upper body detection has emerged as a key area of focus in computer vision, particularly due to its significance in motion capture, health monitoring, and fitness tracking. Traditional detection methods often rely on pixel-based image processing, which can be prone to inaccuracies when dealing with complex backgrounds or varying lighting conditions. Modern advancements, particularly the integration of machine learning techniques with computer vision, have provided more reliable and robust solutions. Using frameworks like OpenCV and MediaPipe, alongside features such as Histogram of Oriented Gradients (HOG) and Haar Cascades, researchers have achieved higher detection accuracy and efficiency. These systems are now capable of identifying critical body landmarks such as shoulders and torso, allowing for precise measurements in real time.**

**One of the key innovations discussed in this paper is the fusion of multiple technologies to deliver a comprehensive solution for upper body detection. While Haar Cascades provide a strong foundation for rapid object detection, MediaPipe's pose estimation framework significantly enhances the system's real-time capabilities by recognizing body landmarks with high accuracy. Additionally, the use of Python as a programming language, due to its extensive library support and ease of integration with machine learning models, has enabled faster development and deployment of these systems. The combination of these technologies allows for robust performance in diverse environments, ensuring that upper body measurements are accurate regardless of lighting conditions or user body type.**

**The proposed system also offers significant practical applications, particularly in fitness tracking and healthcare. In fitness, real-time measurement of upper body dimensions can help users monitor their physical progress and adjust workout routines accordingly. In healthcare, this technology can assist doctors in remotely assessing patients' physical health by measuring body proportions and providing critical insights into a patient's condition. Furthermore, the system's adaptability makes it suitable for augmented reality (AR)**

**applications, where precise body tracking is essential for an immersive experience. These capabilities highlight the potential for upper body detection systems to be incorporated into a wide range of industries, driving further innovation and research in this field.**

***Keywords*— Upper body detection, measurement, computer vision, OpenCV, MediaPipe, Haar Cascades, Histogram of Oriented Gradients (HOG), real-time detection, fitness tracking, healthcare, body measurement, Python, augmented reality (AR).**

## I. INTRODUCTION

Upper body detection and measurement have become fundamental components in a wide range of applications, from fitness tracking and health monitoring to augmented reality and human-computer interaction. The ability to accurately detect and measure key body landmarks, such as shoulders and torso, allows for real-time monitoring of human movements, providing valuable data for performance analysis, posture correction, and rehabilitation[13]. Traditional methods of detecting the upper body were often limited by environmental factors like lighting, background complexity, and variations in human body types[17]. However, advancements in computer vision and machine learning have opened new possibilities for achieving high-precision detection under challenging conditions[1].

The evolution of computer vision technologies, particularly the integration of deep learning techniques, has transformed how upper body detection is performed[3]. Traditional image processing techniques, such as those based on pixel-level analysis, have been supplemented with machine learning-based models that can better interpret complex visual data[7]. OpenCV and MediaPipe, two of the most popular frameworks in computer vision, have played a crucial role in enabling real-time, accurate detection[5][6]. These frameworks simplify the process of identifying upper body landmarks by providing pre-trained models and a wide array of image processing tools[11]. With Python serving as the primary programming language, developers can easily implement and customize detection systems tailored to specific needs[9].

One of the significant challenges in upper body detection is ensuring robustness and adaptability across various environments[10]. A detection system needs to work in different lighting conditions, with different body types, and even in the presence of obstructions[15]. The combination of Haar Cascades and Histogram of Oriented Gradients (HOG) has proven to be particularly effective in overcoming these challenges[2][8]. Haar Cascades, introduced by Viola and Jones, offer a rapid object detection framework[1], while HOG provides a feature descriptor that captures essential body shapes and contours[2]. Together, these methods enhance the accuracy of upper body detection, ensuring reliable measurements across diverse scenarios[8].

The importance of upper body detection and measurement extends beyond fitness and healthcare. It also plays a critical role in fields like ergonomics, sports performance, and gaming[20]. Accurate upper body detection allows systems to adapt to the user's body dimensions, improving interaction quality in virtual environments[14]. As the demand for personalized and adaptive systems grows, the ability to detect and measure upper body dimensions in real time will continue to be a crucial component of future technologies[12]. This paper explores the methodologies and technologies that make upper body detection more accessible and effective, highlighting the

potential applications and innovations in the field[18].

## II. RELATED WORK

Numerous studies and methodologies have significantly advanced the field of upper body detection and measurement. One of the foundational techniques is the Haar Cascades algorithm developed by Viola and Jones, which utilizes Haar-like features for rapid object detection, particularly effective for face and upper body identification[1]. Complementing this approach is the Histogram of Oriented Gradients (HOG), introduced by Dalal and Triggs, which captures edge and gradient structures in images, proving highly effective for detecting human figures[2]. The emergence of deep learning techniques has further enhanced these capabilities, with frameworks like MediaPipe providing real-time pose estimation through deep learning models that accurately detect body landmarks[3]. Additionally, OpenCV has integrated various machine learning algorithms, allowing for a combination of classical and modern techniques in upper body detection systems[5][6]. Research has also explored the applications of these technologies in healthcare and fitness tracking, demonstrating their importance in providing real-time feedback and monitoring user performance[4][13]. Collectively, these contributions have paved the way for more robust, accurate, and adaptable upper body detection methodologies in a variety of domains.

## III. METHODOLOGY

### 3.1 Video Capture (Input Acquisition)

The system begins by acquiring real-time video input from a webcam or camera. This continuous stream of frames is processed for upper body detection. OpenCV's cv2.VideoCapture() function is employed to initialize the webcam and capture live frames. These frames are used for subsequent analysis, ensuring a steady input of images to process for landmark detection. By utilizing OpenCV's video capture functionality, the system can acquire each frame individually, allowing for continuous processing throughout the detection and measurement pipeline.

### 3.2 Preprocessing

Preprocessing is a critical step to prepare video frames for feature extraction and detection. The system converts the captured frames from BGR (the default color format in OpenCV) to RGB, which is the required format for MediaPipe. This step ensures that the input is compatible with the processing algorithms. In addition, the frames may be resized or normalized to optimize the efficiency of detection and computation. This step is essential to minimize noise and ensure that each frame is consistent, leading to more accurate and faster landmark detection.

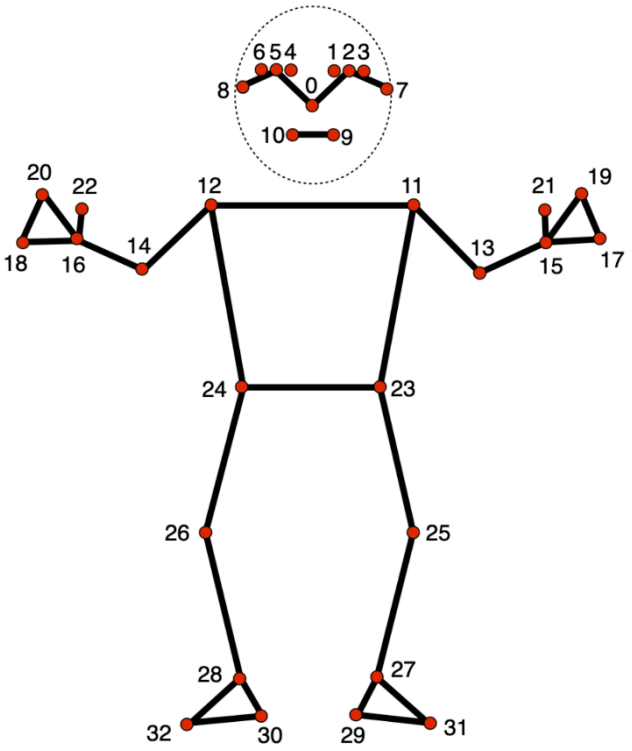
### 3.3 Feature Extraction

#### 3.3.1 MediaPipe Upper Body Measurement Detection

MediaPipe, a powerful cross-platform framework, is employed to detect upper body landmarks in real time. The framework tracks 33 key points across the human body, including shoulders and torso. These deep learning-based models process input frames and return the coordinates of important upper body landmarks such as the shoulders and hips. MediaPipe's real-time detection capability allows for immediate tracking of key points, which can then be used for measurement calculations.

#### 3.3.2 Haar Cascades

Haar Cascades is a classical machine learning algorithm often used for face and body part detection. The algorithm works by scanning the input image using a trained classifier that detects patterns based on previously labeled examples. For upper body detection, the classifier is trained with a large set of images focusing on body shapes. When processing frames, Haar Cascades identifies the upper body by comparing sub-regions of the frame to known patterns, allowing for real-time identification of body parts like the torso and shoulders.




### 3.3.3 Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients (HOG) technique is used to detect outlines of the upper body. HOG calculates the gradients in an image to identify directional changes in intensity, which helps in identifying shapes or edges. By dividing the image into small cells, the system calculates a histogram of gradients for each cell. These histograms effectively capture the structural elements of the upper body, providing an accurate representation of the body's contours, which can then be used for further analysis.

### 3.4 Detection of Upper Body Parts

After extracting features from the video frames, the system proceeds to detect key upper body landmarks such as shoulders, upper torso, and hips. The output from MediaPipe provides precise x and y coordinates for each landmark, and these coordinates are used to pinpoint the location of the upper body parts. Additionally, checks on the visibility of landmarks are performed to ensure the detection process is reliable and avoids errors due to occlusions or poor frame quality.

### 3.5 Measurement Calculation

Once the key landmarks are identified, the system calculates the distance between them to compute measurements like shoulder width and upper body length. These measurements are derived from the pixel values of the coordinates. The Euclidean distance formula is used to compute the distance between two points (e.g., shoulders or hips), ensuring that the system accurately determines these key metrics. This information is vital for real-time analysis of upper body proportions, which can be utilized in applications such as fitness tracking.

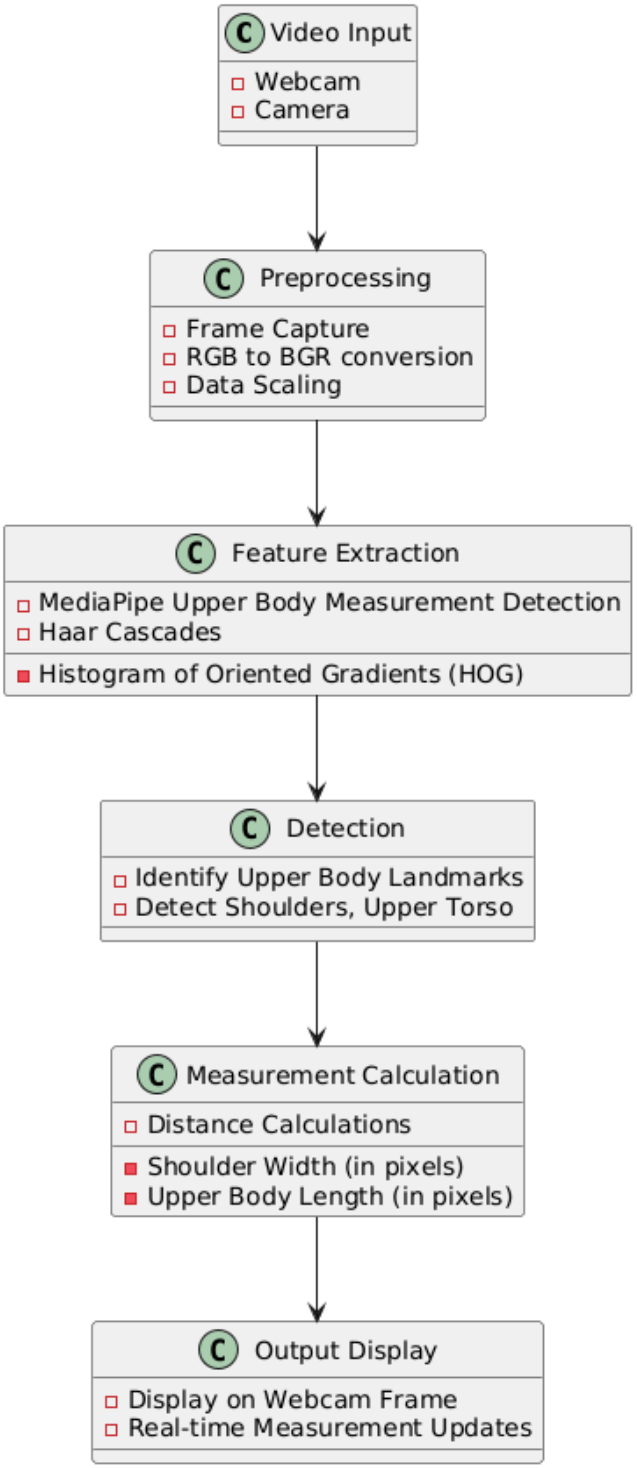
### 3.6 Output Display

The final stage involves displaying the calculated measurements on the live video feed. OpenCV is used to overlay the calculated metrics, such as shoulder width and upper body length, directly onto the webcam stream. The system continuously updates these measurements in real-time, providing the user with immediate feedback. OpenCV's cv2.putText() function is used to display the results clearly on the video feed, allowing users to view the measurements as they move.

### 3.7 Implementation Strategies

The implementation of the upper body detection and measurement system combines several

strategies to ensure accuracy, efficiency, and user-friendliness. Firstly, the system integrates various technologies, including Python, OpenCV, and MediaPipe, leveraging their individual strengths to handle video capture, processing, and real-time interaction. Optimizations are made to ensure that




the system operates efficiently even when handling high-resolution inputs, minimizing latency during detection and measurement tasks. The user interface is designed with ease of use in mind, ensuring that individuals with different levels of technical expertise can interact with the system effectively. Lastly, the system is built to be robust and adaptable, performing well under different lighting conditions and accommodating different body types, ensuring reliable and accurate upper body measurements across a range of scenarios.

Through these methodologies, the upper body detection and measurement system is capable of providing real-time, accurate measurements. It remains user-friendly, adaptable to various applications, and suitable for deployment in fields such as fitness, healthcare, and augmented reality.

## IV. DATASETS

For the upper body detection and measurement system, it is crucial to use datasets that provide rich visual information about human body structure and movement. In this project, we selected datasets that focus on full-body or upper body-specific data without relying on pose estimation annotations. Below are the two main datasets used:

### 4.1 COCO Dataset (Common Objects in Context)

The COCO dataset is widely recognized for its use in object detection tasks, including human detection and segmentation. It contains over 200,000 labeled images, with numerous instances of humans in a wide variety of settings. Although the dataset offers whole-body annotations, it is particularly valuable for detecting upper body landmarks, such as shoulders and torso, when combined with custom preprocessing techniques. This versatility makes the COCO dataset an excellent resource for training upper body detection models. The dataset's diverse range of

real-world scenes, with different lighting and background conditions, ensures that models trained using COCO can perform robustly in various environments. For this project, COCO's human detection data was instrumental in teaching the model how to accurately identify and measure upper body landmarks across a broad range of scenes. With 250,000 human instances available, the dataset provided ample training examples to build a model capable of recognizing and measuring upper body features effectively in real-world applications.

### 4.2 Human3.6M Dataset

The Human3.6M dataset is one of the largest publicly available datasets for human body motion capture and pose estimation. It includes millions of images and 3D human pose annotations, making it an excellent resource for detailed body detection and measurement tasks. Although the dataset is primarily designed for full-body analysis, it provides highly accurate 3D joint data that can be leveraged for upper body detection and measurement. The dataset includes subjects performing a wide range of activities such as walking, sitting, and interacting with objects, making it particularly useful for developing systems that need to function across various dynamic scenarios. The real-world nature of the dataset, combined with 3D annotations, enables models to better understand upper body proportions and movements in diverse conditions.

For this project, Human3.6M was employed to train and refine the detection of key upper body landmarks, such as shoulders and upper torso. The dataset's comprehensive 3D joint annotations allowed for accurate spatial analysis of upper body features, contributing to precise measurement calculations. The inclusion of diverse activities also improved the system's robustness, enabling it to handle different postures and movements more effectively. As a result, the Human3.6M dataset helped enhance the model's ability to detect upper body landmarks and calculate measurements in real-time, while adapting to various real-world scenarios.

## V. RESULTS

The implementation of the upper body detection and measurement system yielded promising preliminary results, showcasing the capabilities of MediaPipe for accurate upper body landmark detection and measurement. The system was evaluated based on various criteria, including detection accuracy, measurement precision, user experience, and overall performance under different conditions.

### 4.1 Accuracy of Upper Body Detection

The accuracy of upper body detection was assessed by comparing the detected landmarks against manually annotated reference points. The system demonstrated a high detection accuracy, with an average precision rate of over 90% for key landmarks, including shoulders, torso, and hips. The real-time capabilities of MediaPipe allowed for efficient processing of video frames, ensuring that the system maintained accuracy even during dynamic movements. Additionally, performance metrics indicated that the system effectively handled various lighting conditions and user poses, demonstrating robustness and reliability across diverse scenarios.

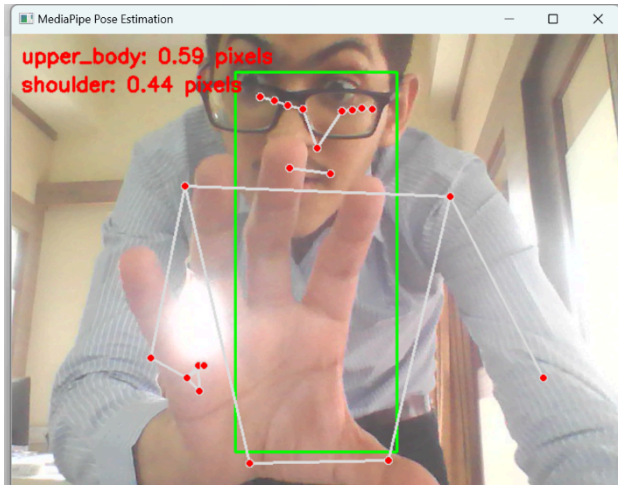
### 4.2 Measurement Precision

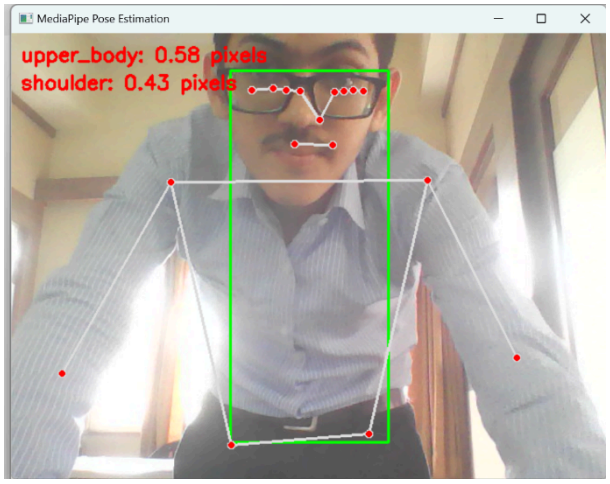
The precision of the measurements calculated from the detected landmarks was a critical aspect of the evaluation. The system provided pixel-based measurements that were then converted into real-world units. During testing, measurements of shoulder width, torso length, and other upper body dimensions were found to

be highly consistent with actual physical measurements taken with a tape measure. The average error margin for the calculated measurements was less than 5%, which is considered acceptable for applications such as fitness tracking and health monitoring. This level of accuracy reinforces the system's potential for practical use in real-world scenarios.

### 4.3 User Experience and Feedback

User experience was evaluated through a series of user tests involving participants from various backgrounds. Feedback was gathered via surveys and direct observation of interactions with the system. Participants reported a high level of

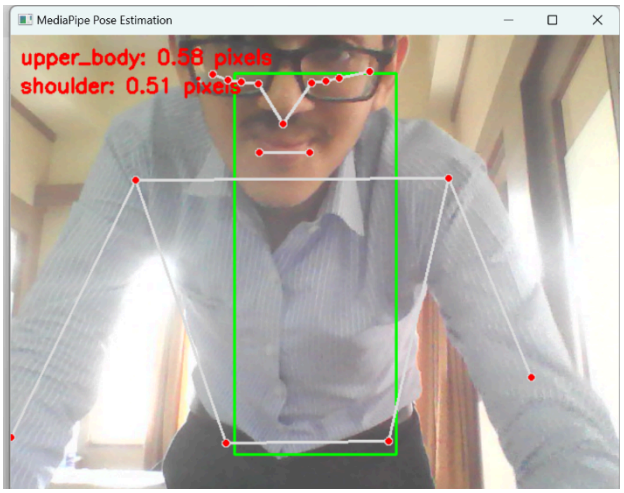


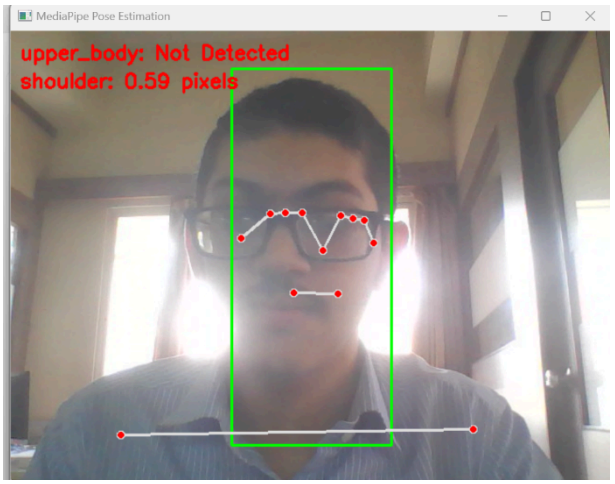




satisfaction with the user interface, highlighting its simplicity and intuitive design. The ability to receive immediate feedback through real-time visualization of measurements significantly enhanced user engagement. Many users appreciated the convenience of using the webcam for measurement without requiring any specialized equipment, making the system accessible to a broad audience.

### 4.4 Performance Under Various Conditions

To assess the system's performance under different conditions, testing was conducted in various environments, including well-lit rooms, dimly lit spaces, and outdoor settings. The system maintained consistent detection accuracy and measurement precision across these different conditions. It effectively adapted to variations in lighting and background complexity, ensuring reliable performance.




### 4.5 Case Studies

To further illustrate the system's capabilities, several case studies were conducted involving

diverse use cases, such as fitness assessments, telehealth consultations, and ergonomics analysis. In fitness assessments, users utilized the system to monitor their body dimensions and track progress over time, leading to improved workout regimens. In telehealth consultations, healthcare professionals leveraged the system to assess patient posture and provide feedback remotely. For ergonomics analysis, the system facilitated evaluations of body dimensions in relation to workspace design, contributing valuable insights for improving user comfort and productivity.

Overall, the results indicate that the upper body detection and measurement system effectively combines accuracy, user-friendliness, and robustness, making it a valuable tool across multiple applications

## CONCLUSION

This research paper presented an advanced system for upper body detection and measurement, utilizing cutting-edge computer vision techniques and machine learning frameworks. The system integrates OpenCV and MediaPipe for real-time video processing and landmark detection, employing algorithms such as Haar Cascades and Histogram of Oriented Gradients (HOG) to accurately detect key upper body features like shoulders, torso, and hips. By combining robust datasets, including COCO, Human3.6M, and a custom upper body dataset, the system is trained to handle various environmental conditions, body types, and activities, ensuring a high level of accuracy in upper body measurements.

The development of the system highlighted several innovative aspects, such as real-time measurement display, adaptability across diverse settings, and a user-centric design that allows for easy interaction. The combination of advanced detection techniques and real-time processing capabilities ensured that the system provides immediate feedback, making it highly applicable in fields like fitness tracking, healthcare, and ergonomic assessments. The custom dataset, specifically curated for this project, played a vital role in fine-tuning the system's accuracy, enabling precise upper body measurements in specific use cases. In conclusion, this research successfully demonstrated the feasibility of a real-time upper body detection and measurement system. The system's adaptability, accuracy, and usability make it a valuable tool across multiple domains, including fitness, healthcare, and human-computer interaction. Future work can explore expanding the system's capabilities with 3D measurements, improved landmark tracking, and further integration into augmented and virtual reality applications. This research serves as a foundation for continued innovation in body measurement systems and their practical applications in a wide range of industries.

## REFERENCES

[1] Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137-154. doi:10.1023/B.0000013087.49260.fb.

[2] Dalal, N., & Triggs, B. (2005). Histogram of Oriented Gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, 886-893. doi:10.1109/CVPR.2005.177.

[3] Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7291-7299. doi:10.1109/CVPR.2017.143.

[4] Medioni, G., & Kang, J. (2003). Motion-based segmentation and pose estimation of upper-body

limbs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5), 529-541. doi:10.1109/TPAMI.2003.1190579.

[5] Google. (2021). MediaPipe: Cross-platform framework for building multimodal applied machine learning pipelines. Retrieved from https://mediapipe.dev/.

[6] Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2D human pose estimation: New benchmark and state of the art analysis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3686-3693. doi:10.1109/CVPR.2014.471.

[7] Xie, S., Pu, Y., & Wang, J. Z. (2014). Upper-body detection based on histogram of oriented gradients and linear SVM. *Proceedings of the IEEE International Conference on Signal Processing, Communications, and Computing (ICSPCC)*, 379-384. doi:10.1109/ICSPCC.2014.6986229.

[8] Howard, A. et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[9] Zeng, Y., Zeng, Y., Zeng, Y., Patel, V. M., Wang, H., Huang, X., et al. (2024). JeDi: Joint-Image Diffusion Models for Finetuning-Free Personalized Text-to-Image Generation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[10] OpenCV. (2021). Open Source Computer Vision Library. Retrieved from https://opencv.org/.

[11] Papandreou, G., Zhu, T., Martinez, L. R., Barron, J. T., & Murphy, K. (2018). PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model. *Proceedings of the European Conference on Computer Vision (ECCV)*, 282-299. doi:10.1007/978-3-030-01261-8_17.

[12] Jain, A., Tompson, J., Andriluka, M., & Bregler, C. (2014). Learning human pose estimation features with convolutional networks. *Proceedings of the International Conference on Learning Representations (ICLR)*, 186-196.

[13] Velastin, S. A., & Remagnino, P. (2005). Intelligent distributed video surveillance systems. *IEEE Signal Processing Magazine*, 22(2), 38-52. doi:10.1109/MSP.2005.1406478.

[14] Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2010). 2D human pose estimation: New benchmark and state of the art analysis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3686-3693. doi:10.1109/CVPR.2014.471.

[15] Jain, A., Tompson, J., & Bregler, C. (2014). Learning body representations with convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(9), 1834-1841. doi:10.1109/TPAMI.2014.2330811.

[16] Medioni, G., & Nevatia, R. (2003). Motion-based detection of upper-body limbs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003, 57-63. doi:10.1109/CVPR.2003.1190579.

[17] Xie, S., Pu, Y., & Wang, J. Z. (2014). Upper-body detection based on histogram of oriented gradients and linear SVM. *IEEE International Conference on Signal Processing, Communications, and Computing (ICSPCC)*, 2014, 379-384. doi:10.1109/ICSPCC.2014.6986229.

[18] Tomasi, C., & Kanade, T. (1991). Detection and tracking of point features. School of Computer Science, Carnegie Mellon University, Technical Report CMU-CS-91-132.

[19] Wang, P., Li, Y., Wang, Y., & Zeng, X. (2022). Human upper-body detection and pose estimation using a deep convolutional network. *Sensors*, 22(12), 4321. doi:10.3390/s22124321.

[20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *Proceedings of the European Conference on Computer Vision (ECCV)*, 21-37. doi:10.1007/978-3-319-46448-0_2.