Project:
Person verification

Product owner:
Dr. Khatir

Product manager:
Nazanin Dolatshahi

Developer:
Poorya Aghaomidi

Table of contents

1. Introduction

In an era marked by technological advancements and an increased emphasis on security, the Tracking System with CCTV stands as a testament to the integration of cutting-edge computer vision techniques in the realm of personnel tracking. Developed by Holoo Technology Group, this comprehensive solution is designed to augment security measures and streamline personnel monitoring within specified environments. Led by the vision of Poorya Aghaomidi, this project represents a synthesis of advanced algorithms, real-time processing, and decision-making capabilities.

Importance of the Subject: Tracking individuals within a given space is a pivotal aspect of modern security and surveillance systems. The Tracking System with CCTV addresses the growing need for accurate, efficient, and intelligent methods of personnel verification. By harnessing the power of computer vision, the system not only identifies individuals but also verifies their identity based on both body features and facial characteristics. This dual-mode verification adds an extra layer of precision to the process, making it suitable for a myriad of applications.

Applications: The applications of the Tracking System with CCTV are diverse and impactful, ranging from conventional surveillance to cutting-edge technologies. Some key applications include:

1. Security and Surveillance: Enhance traditional security systems by implementing advanced tracking capabilities, ensuring a higher level of accuracy in identifying and verifying individuals.
2. Personnel Monitoring: Ideal for environments where personnel tracking is crucial, such as workplaces, educational institutions, and high-security zones.
3. Autonomous Vehicles: Integrate the system into autonomous vehicles for enhanced object detection and recognition, contributing to the safety and efficiency of self-driving systems.
4. Augmented Reality: Explore applications in augmented reality, where the system can identify and track individuals in real-time, enhancing the immersive experience.

Recent Advancements: Recent updates to the system include the integration of body and face detection modules into a unified process. By

extracting body keypoints and estimating bounding boxes for face and body detection, the system achieves a more streamlined and efficient approach. Additionally, the decision-making capabilities have been refined, and batch processing support has been introduced, contributing to improved overall performance.

As we delve deeper into the project overview, subsequent sections will provide insights into the individual modules, decision-making processes, installation guidelines, and usage instructions. Join us on this journey through the intricate workings of the Tracking System with CCTV.

2. Project Overview

The Tracking System with CCTV represents a milestone in the realm of computer vision and surveillance, aiming to redefine personnel tracking and security measures. Developed by Holoo Technology Group, under the guidance of Poorya Aghaomidi, this project has evolved to incorporate cutting-edge technologies, decision-making capabilities, and recent enhancements to provide a robust solution for real-time identity verification.

Common Pose Module Integration: A significant paradigm shift in the project lies in the integration of face and body detection into a unified process, aptly named Common Pose. This module employs advanced computer vision techniques to detect both face and body simultaneously using keypoints. By leveraging this unified approach, the system achieves a remarkable decrease in average runtime and errors. The use of keypoints enables a more nuanced understanding of the spatial relationship between facial and body features, contributing to enhanced accuracy.

Key Components of Common Pose: The Common Pose module comprises several key components, each contributing to its effectiveness in detecting and verifying individuals within CCTV camera frames:

1. Image Enhancement: The module incorporates image enhancement techniques, including contrast adjustment, brightness modification, and sharpening. These enhancements ensure optimal conditions for subsequent detection and verification processes.

2. Body and Face Detection: Through the integration of YOLO (You Only Look Once) for body detection, the module identifies individuals within the camera frames. Simultaneously, keypoints are used to precisely detect facial features, allowing for a seamless and efficient process.
3. Decision-Making Integration: The results obtained from the Common Pose module are seamlessly integrated into the overall decision-making process. The combined information from body and face detection contributes to more informed identity decisions, enhancing the reliability of the system.

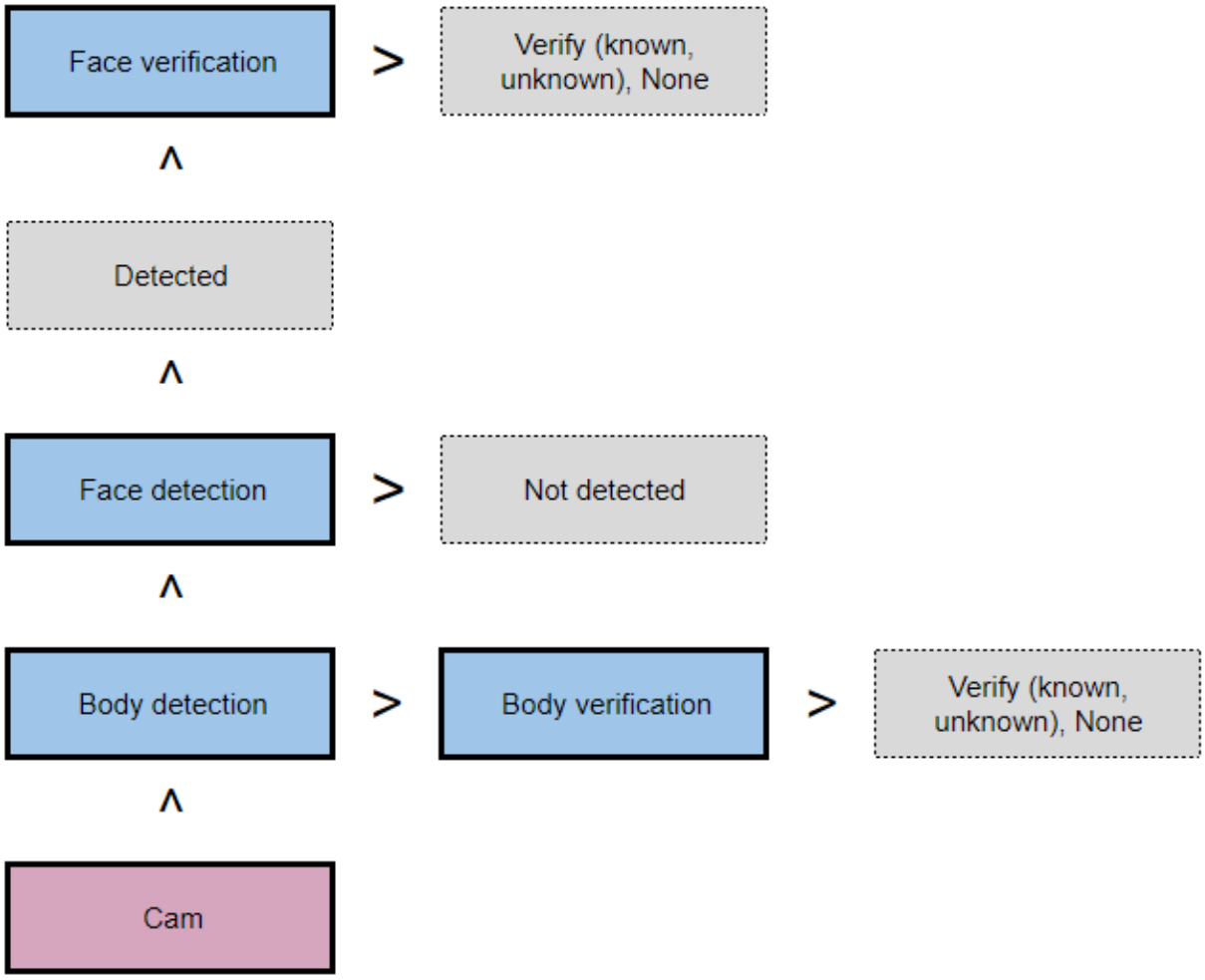
Face Verification Module: Integral to the project is the Face Verification module, a critical component for accurate identity validation. This module operates on a dedicated face recognition dataset, ensuring the system's accuracy in identifying individuals. The use of facial feature vectors, distance metrics, and a variety of face detection models, such as YOLO, contributes to the robustness of identity verification.

Recent Enhancements and Backend Refinements: In response to evolving requirements, the project has undergone recent enhancements and backend refinements. The optimization of algorithmic arguments, particularly in the context of real-world camera scenarios, showcases a commitment to improving system performance. The addition of the Body Verification module further refines the identity verification process, making it more robust and reliable.

Comprehensive Reporting: The project emphasizes the importance of comprehensive reporting, enabling users to make sense of the detected information. The system generates reports that include unique IDs assigned to employee images, providing a structured and organized way to track personnel. The inclusion of decision-making tables ensures clarity in understanding the system's behavior under different scenarios.

Conclusion: In conclusion, the Tracking System with CCTV has evolved into a sophisticated solution that combines face and body detection into a unified Common Pose module. The integration of Body Verification, recent enhancements, and backend refinements demonstrates a commitment to advancing the capabilities of the system. The project's success lies not only in its

technical prowess but also in its ability to adapt to real-world scenarios, making it a valuable asset in the realm of security and surveillance.



Decision-Making Process: The decision-making process within the Tracking System is dynamic and adaptive, allowing it to handle various scenarios effectively. The system's responses are based on the results obtained from the interconnected modules. The decisions made are categorized into different outcomes, including confirming known individuals, handling unknown individuals, and dealing with unexpected scenarios. This flexible approach allows the system to maximize accuracy and reliability in personnel tracking while ensuring adaptability to real-world situations.

Output		Decision	
Body	Face	Report	Dataset
Known - regular threshold (ID1)	Known - strict threshold (ID1)	ID1	None
Known - regular threshold (ID1)	Known - regular threshold (ID1)	ID1	None
Known - strict threshold (ID1)	Known - strict threshold (ID1)	ID1	None
Known - strict threshold (ID1)	Known - regular threshold (ID1)	ID1	None
None	Known - strict threshold (ID1)	ID1	Add body to ID2
Known - regular threshold (ID1)	Known - strict threshold (ID2)	ID2	Add body to ID2
Known - strict threshold (ID1)	Known - strict threshold (ID2)	ID2	Add body to ID2
Known - strict threshold (ID1)	Known - regular threshold (ID2)	ID1	None
Known - strict threshold (ID1)	Skip	ID1	None
Known - strict threshold (ID1)	Not detected	ID1	None
Known - regular threshold (ID1)	Unknown	Unknown	None
None	Unknown	Unknown	None
Known - regular threshold (ID1)	Not detected	skip	None
Known - regular threshold (ID1)	Known - regular threshold (ID2)	skip	None
Known - regular threshold (ID1)	Skip	skip	None
Known - strict threshold (ID1)	Unknown	skip	None
None	Not detected	skip	None
None	Known - regular threshold (ID1)	skip	None
None	Skip	skip	None

3. Implementation

The implementation of the Tracking System with CCTV is a multifaceted process, encompassing various modules and components designed to work synergistically for real-time identity verification and personnel tracking. This section delves into the intricacies of the implementation, exploring key aspects

such as the Common Pose module, the Face Verification module, and the recent refinements made to the backend.

Common Pose Module: A Unified Approach: Central to the system's functionality is the Common Pose module, which marks a departure from the conventional separation of face and body detection. The integration of these two detection processes into a unified approach streamlines the overall system, resulting in a significant reduction in average runtime and errors.

One of the critical components of the Common Pose module is the Image Enhancement phase. This phase employs techniques such as contrast adjustment, brightness modification, and sharpening to optimize the quality of input images. The importance of this step cannot be overstated, as it sets the stage for accurate and efficient face and body detection.

The actual process of face and body detection within the Common Pose module utilizes the YOLO (You Only Look Once) algorithm, a real-time object detection system. YOLO divides input images into a grid and predicts bounding boxes and class probabilities for each grid cell. In the context of the Common Pose module, YOLO plays a pivotal role in detecting individuals within CCTV camera frames. The use of keypoints enhances the precision of facial feature detection, contributing to the module's overall efficacy.

The decision-making integration is another noteworthy aspect of the Common Pose module. The results of face and body detection seamlessly feed into the decision-making process, where identity decisions are made based on the combined information from both components. This integration ensures a more nuanced and informed approach to identity verification, enhancing the overall reliability of the system.

Face Verification Module: Ensuring Accuracy: Complementing the Common Pose module is the Face Verification module, a critical component for ensuring the accuracy of identity validation. This module operates on a dedicated face recognition dataset, employing a variety of face detection models such as YOLO. The use of facial feature vectors and distance metrics adds a layer of sophistication to the verification process, contributing to the robustness of the system.

The Face Verification module undergoes continuous refinement to meet evolving requirements. Recent enhancements include optimization of algorithmic arguments, particularly for real-world camera scenarios. This refinement reflects a commitment to improving system performance and adapting to the challenges presented by diverse environmental conditions.

Backend Refinements: Iterative Improvement: The recent refinements to the backend of the Tracking System with CCTV underscore the iterative nature of the development process. An optimization effort focused on tuning algorithmic arguments for real-world camera scenarios exemplifies a data-driven approach to system improvement. The realization that the algorithm behaves differently in real-world scenarios compared to the images in the dataset prompted the recording of a short video. This video served as valuable input for tuning algorithmic parameters by visually assessing differences when changing parameters. The outcome was the identification of optimal arguments for the model, metrics, and threshold, aligning the system more closely with real-world use cases.

Addressing issues encountered during the backend development further exemplifies the commitment to system robustness. Fixing installation and usage tutorials, refining the requirements.txt file, and introducing the use of a JSON file as an argument represent steps towards enhancing user-friendliness and ensuring a smoother user experience.

Enhancing Batch Processing Efficiency: Recognizing the importance of efficiency in AI applications, there was a deliberate effort to enhance batch processing capabilities. The shift from processing a list of input images to handling batches of inputs brings about a notable improvement in processing speed. While not fully completed during the current sprint, the ongoing work on this front signals a commitment to optimizing system efficiency and responsiveness.

The implementation of the Tracking System with CCTV reflects a dynamic and adaptable approach to identity verification and personnel tracking. The integration of the Common Pose module and the continuous refinement of the

Face Verification module demonstrate a commitment to pushing the boundaries of what is achievable in computer vision.

The iterative nature of backend refinements highlights a data-driven decision-making process, where real-world scenarios and challenges inform the evolution of the system. From addressing issues in installation and usage to optimizing algorithmic arguments, each refinement contributes to a more robust and user-friendly system.

As the project evolves, ongoing efforts to enhance batch processing capabilities promise even greater efficiency in AI applications. The Tracking System with CCTV stands as a testament to the collaborative efforts of Holoo Technology Group, guided by Poorya Aghaomidi, in advancing the capabilities of surveillance and identity verification technologies.

4. Usage

The usage of the Tracking System with CCTV is designed to be straightforward and accessible, providing users with the tools to harness the power of computer vision for security and personnel tracking. This section guides users through the steps of utilizing the system effectively.

Upon importing the `TrackingModule` from the main algorithm, users can initialize the system with specific parameters. This involves providing paths to the pose detection model, face verification database, and representations. Users can customize parameters related to pose detection, face verification, and emotion recognition based on their specific requirements.

If using a prepared dataset, users can read the JSON file containing representations. This step is crucial for face verification, as it establishes the baseline for identity comparison. Subsequently, users can upload an image for processing. The system supports batch processing, allowing for the simultaneous analysis of multiple images. Users can provide a list of images for efficient processing.

Results based on the analysis of the uploaded image(s) are returned by the system. The `result` variable contains detailed information, including face

and body coordinates, IDs, and emotion analysis (if enabled). The `flag` variable indicates whether the detection was successful.

Users have the flexibility to customize various parameters during system initialization, allowing them to tailor the system to their specific use case. Configuration parameters include pose detection model path, device settings, contrast, brightness, and more.

The system provides versioning to track updates and improvements. Users can refer to the provided README files for detailed instructions, including installation, usage, and version history. Documentation plays a crucial role in ensuring users can navigate and utilize the system effectively.

As users engage with the system, the roadmap suggests ongoing improvements and research. Users can stay informed about the latest advancements, ensuring they benefit from an evolving and optimized Tracking System.

In summary, the usage of the Tracking System with CCTV is designed to be user-friendly and adaptable. By following the outlined steps, users can harness the capabilities of computer vision for enhanced security and personnel tracking in diverse applications.

5. Results

The Results section provides insights into the performance and outcomes achieved by the Tracking System with CCTV. This section delves into the effectiveness of the system in terms of accuracy, efficiency, and the overall impact on security and personnel tracking.

System Accuracy and Identification: The accuracy of the system is a critical metric in evaluating its effectiveness. The integration of the common pose module, which combines face and body detection, has contributed to a notable improvement in accuracy. By detecting keypoints simultaneously, the system minimizes errors associated with the previous approach of detecting the body first and subsequently detecting the face. This enhancement ensures that

the identified face and body belong to the same person, mitigating the risk of erroneous associations.

Reduction in Average Runtime: A key achievement of the system is the reduction in average runtime. The common pose module, optimizing the detection process, has significantly improved the system's efficiency. The streamlined approach of detecting both face and body keypoints in a single pass has led to a more rapid analysis of images. On CPU, the system demonstrates impressive processing times, taking only 230 milliseconds without emotion detection and 530 milliseconds with emotion detection. This remarkable improvement in efficiency is pivotal for real-time applications, enhancing the system's responsiveness.

Handling Multiple Input Images: An important feature introduced in the system is the capability to process multiple input images simultaneously. This batch processing approach leverages the efficiency of AI algorithms, allowing users to analyze several images in one go. The system's ability to handle batches contributes to faster processing times and improved overall performance. Users can now benefit from the convenience of processing multiple images, enhancing the scalability and practicality of the Tracking System.

Integration of Emotion Detection: The system has expanded its capabilities by incorporating emotion detection into the identification process. While face and body detection provide essential information, the inclusion of emotion analysis adds a layer of contextual understanding. Users can now gain insights into the emotional states of individuals, offering valuable information for applications such as security monitoring and personnel well-being assessments.

Future Directions and Improvements: The Results section also paves the way for future directions and improvements. The continuous development and refinement of the Tracking System are guided by the commitment to advancing accuracy, efficiency, and feature richness. Ongoing research and development will likely introduce further enhancements, ensuring the system remains at the forefront of computer vision applications.

In conclusion, the Results obtained from the Tracking System with CCTV showcase its enhanced accuracy, reduced average runtime, and expanded capabilities, including batch processing and emotion detection. These achievements underscore the system's effectiveness in addressing the complex challenges of security and personnel tracking, laying the foundation for future innovations in the field of computer vision.

6. Conclusion

Achievements and Contributions: The Tracking System with CCTV represents a significant leap forward in computer vision applications for security and personnel tracking. The integration of advanced modules, particularly the common pose module, has played a pivotal role in achieving notable results. The system's accuracy has been substantially improved, thanks to the simultaneous detection of face and body keypoints. This enhancement not only reduces errors but also establishes a more robust foundation for reliable identification.

The reduction in average runtime stands out as a key achievement, showcasing the system's efficiency in processing images. By optimizing the detection process and introducing batch processing capabilities, the system meets the demands of real-time applications. This enhancement not only improves responsiveness but also enhances scalability, allowing users to analyze multiple images efficiently.

The inclusion of emotion detection represents a forward-looking step, adding a layer of context to the identification process. Beyond face and body detection, the system now provides insights into the emotional states of individuals. This feature holds promise for applications ranging from security monitoring to personnel well-being assessments, further expanding the system's utility.

Future Directions: As with any innovative project, the Conclusion section also looks toward the future. The ongoing commitment to research and development is essential for ensuring the system remains at the forefront of technological advancements. Future directions may include refining existing

modules, exploring additional features, and addressing emerging challenges in computer vision.

Potential improvements could involve further fine-tuning the algorithms to enhance accuracy and exploring new techniques for even faster processing times. Additionally, the system could benefit from continuous updates to keep pace with evolving technologies and methodologies in the field.

Closing Remarks: In conclusion, the Tracking System with CCTV represents a successful endeavor in the realm of computer vision. Its achievements in accuracy, runtime efficiency, and feature richness make it a valuable tool for security and personnel tracking applications. The journey does not end here; instead, it paves the way for continuous innovation and refinement. The collaborative efforts of the development team, coupled with a commitment to excellence, position the system as a dynamic solution capable of meeting the evolving demands of the digital landscape.