The Republic of Iraq

Ministry of Higher Education and Scientific Research

University of Babylon \\College of Science for Girls

Department of Computer Science

Research Title

(Face Detection For Person Tracking)

Research submitted to the Department of Computer Science at the College of Science for Girls at the University of Babylon It is one of the requirements for obtaining a bachelor's degree in computer science

Done by

Shaima Jassim Mohammed

Supervised by

Dr. Salah Mahdi Al-Obaidi

2023-2024

اقرار المشرف

اشهد ان اعداد هذا المشروع بعنوان (Face Detection For Person Tracking) قد جرى في قسم علوم الحاسوب في كلية العلوم للبنات / جامعة بابل , وهو جزء من متطلبات نيل شهادة البكالوريوس في علوم الحاسبات من قبل طالبة المرحلة الرابعة (شيماء جاسم محمد)

توقيع المشرف

اسم المشرف:- صلاح مهدي العبيدي

المرتبة العلمية:- دكتور

التاريخ: 2023-2024

Abstract

In the world of computer vision, face detection is a fundamental and exciting task. Detecting and locating faces in images or video streams forms the basis of many applications, from face recognition systems to digital image processing. Among the many algorithms that have been developed to deal with this challenge, the Viola-Jones algorithm has emerged as a pioneering approach famous for its speed and accuracy. Developed by Paul Viola and Michael Jones in 2001, the Viola-Jones algorithm changed the field of face detection. Its efficient and powerful methodology has opened doors to a wide range of applications that rely on accurately identifying and analyzing human faces. By exploiting the power of hair-like features, whole images, machine learning, and cascade classifications, the Viola-Jones algorithm demonstrates the synergy between computer science and image processing.

In this project, we use face detection technology and track people's movement using the Viola Jones algorithm, with a focus on identifying danger zones to monitor people's traffic. This technology is distinguished by its ability to recognize faces with high accuracy and track people in real time. By processing and analyzing images, potential areas of danger are identified and alerted if someone passes through them.

بِسْمِ اللَّـهِ الرَّحْمَـٰنِ الرَّحِيمِ

﴿يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ﴾

صدق الله العلي العظيم

المجادلة :﴿11﴾

الاهداء

إلى أمي الحنونة، الشمعة التي تنير دربي برعايتها الدائمة والتي كان لها فضل كبير في وصولي إلى هنا. إلى أبي الغالي، الركيزة التي تسندني في حياتي. وإلى بقية العائلة، الذين يشكلون دعامة قوية وجزءًا من حياتي.

الشكر والتقدير

أتقدم بخالص الشكر والتقدير للمشرف على هذا البحث (دكتور صلاح مهدي العبيدي)، على توجيهاته القيمة ودعمه الذي ساهم بشكل كبير في نجاح هذا البحث. بفضل مساهمته الحكيمة والخبرة التي قدمها، نجحنا في تحقيق نتائج موثوقة وذات جودة عالية.

كما أود أن أعرب عن خالص امتناني وشكري لجميع الأفراد الذين ساهموا في تقدمي وتطوري خلال المسيرة الدراسية. دعمهم وتحفيزهم كانا دافعًا قويًا لتحقيق النجاحات والتطورات التي حققناها. شكرًا لكم جميعًا على تعاونكم ومساهماتكم القيمة.

Table of contents

Chapter One
Introduction

1-1 Introduction

Face detection for people tracking is a fundamental problem in the field of computer vision and image processing. This research aims to review the methods and techniques used in face detection and people tracking without relying on deep learning. Face detection technology for tracking people on video is important in many practical applications such as face recognition, security surveillance, and machine learning applications. Face detection is an essential first step in face verification , face identification and clustering , facial landmarks , facial hallmark classification , face alignment , and face tracking [1]. There are several ways to solve the problem of face detection and tracking due to its advantages and limitations, and some of the methods include:

1- Viola-Jones method:- This learning-based algorithm is used for object detection. It employs Haar features and a cascade of classifiers to identify objects. The Haar features are computed using the integral image. The Adaptive Boost (Adaboost) algorithm is then used to select the best features [2].

2- Feature-based methods:- Feature based face recognition uses a priori information or local features of faces to select a number of features to uniquely identify individuals. Local features include the eyes, nose, mouth, chin and head outline, which are selected from face images [3].

3- Deep learning methods:- The technique that was used in the development of a face recognition system. It was used a deep learning approach where the CNN model was developed and trained by using the different number of images [4].

4- Motion-based methods:- A technique for identifying subtle movements in several areas of the face using the motion detection method. There are two motion detection methods: optical flow method and block matching algorithm (BMA) method. The optical flow method uses the Kanade-Lucas Tomasi (KLT) method and the BMA method uses the phase-only correlation (POC) algorithm [5].

Similar to the general object tracking problem, face tracking also has challenges such as losing the target when there are appearance changes or occlusions; computational efficiency issues when there are multiple targets; difficulty in performing data association across multiple cameras and so on. Moreover, there are additional obstacles specific to face tracking. For instance, the tracking model should be robust to quick motions of the human head ,and it should be able to keep tracking a face even when the face turns away from the camera and then comes back. Furthermore, as tracked faces often need to be identified, how to select representative faces from a face sequence for optimal identification performance is another critical research component [6]. It is certain that there are several solutions to meet detection challenges, including. Improving performance in low-light conditions, enhancing facial recognition in non-standard situations, using 3D facial recognition techniques, taking advantage of deep learning for facial recognition, applying advanced algorithms, and using multimodal methods.

Based on previous investigations in the field of face detection for people tracking, this paper will examine the Viola-Jones method of face detection for people tracking in video and some video enhancement techniques used to improve video quality and accuracy of face detection in consecutive video frames will also be studied.

1-2 State-Of-The-Art face Tracking Methods

There are different ways to categorize face tracking algorithms. For example, it might be based on whether the entire face is tracked as a single entity or individual facial features are tracked; or, it might be based on whether the tracking is in the 2D image space or in the 3D pose space [6].

1-2-1 The KLT method

The Kanade–Lucas–Tomasi (KLT) feature tracker is a classic approach for object tracking which makes use of spatial intensity information to direct the search for the position of a neighborhood in a sequential frame that yields the best match. Built upon three basic assumptions, i.e. brightness constancy (the brightness of a pixel does not change from frame to frame), temporal persistence (the object does not move much from frame to frame) and spatial coherence (neighboring points have similar motion),the LK algorithm relies only on local information that is derived from some small window surrounding each of the points of interest, The KLT algorithm has been used as the core tracking method in many existing face tracking approaches.

The basic idea is to use the KLT tracker to track points of interest throughout video frames, and then each face track is formed by the faces detected in different frames that share a large enough number of tracked point. The Open CV implementation of Viola-Jones face detector is applied to every frame. Detected faces are added into face tracks based on the grouping of tracked points.

1-2-2 The Mean Shift / Cam Shift methods

The Mean Shift is a general data analysis method with many applications, and computer vision is only one of them. It is a non-parametric feature-space analysis technique for estimating density gradients. Comaniciu et al [7]. used the mean shift algorithm to perform object tracking in feature space where mean shift iterations are conducted to find the most probable target position in the current frame. The dissimilarity between the target model (described by its color distribution in an intensity normalized space) and the target candidates is expressed by a distance metric. It provides a practical, fast and efficient solution for tracking objects.

The Cam Shift (Continuously Adaptive Mean Shift) algorithm extends Mean Shift with an adaptive region-sizing step, which is more appropriate for tracking objects whose size may change during a video sequence [8]. That is, after mean shift is run, the size and angle of the target window are estimated and updated.

The primary difference between the Cam Shift and the Mean Shift is that Cam Shift uses continuously adaptive probability distributions (i.e. distributions that may be updated for each frame); while Mean Shift uses static distributions which are not updated unless there are significant changes in shape, size or color of the target [9].

1-2-3 The Particle Filter methods

Particle filters are sequential Monte Carlo methods based on point mass (or "particle") representations of probability densities, can be applied to any state-space model. The goal of a particle filter-based tracker is to maintain a probability distribution over the state (location, scale, etc.) of tracked objects. This distribution is normally represented as a set of weighted particles (or samples, which represent possible locations of the tracked object). This weighted distribution is updated over time by means of a set of Bayesian filtering equations. The particle with the highest weight or the weighted mean of the particle set at each time step defines the location of the object. the particle filter method has the advantage of recovering, to a certain extent, from a loss of track in the scenarios of occlusion and clutter.

According to [10]. there are two major types of visual trackers: the "Target Representation and Localization" ,The particle filter belongs to the second type. It is robust to occlusions and can deal with multi-modal PDFs.

1-2-4 The TLD method

Recent tracking approaches employ an adaptive tracking-by-detection strategy, in which a detector predicts the position of an object and adapts its parameters to the current object's appearance. In the TLD framework proposed by Kalal et al [11] . the long-term tracking task (i.e. the process should run indefinitely long) is decomposed into three sub-tasks: tracking, learning and detection.

They invented the P-N learning paradigm that estimates detection errors and uses these errors to bootstrap the classifier. In every frame, the P-N learning performs the following steps: (1) evaluation of the detector on the current frame, (2) estimation of detector errors using the P-N experts, (3) update of the detector by labeled examples output by the experts. While in the original TLD, the entire detector is learned online, in contrast to the original TLD, the detector is split into two parts: an offline-trained generic face detector that localizes frontal faces and an online-trained validator that decides which faces correspond to the tracked subject.

1-2-5 The PHD filters for multiple target tracking

The PHD (Probability Hypothesis Density) filter is a multiple-target filter for recursively estimating the number and the state of a set of targets given a set of observations. It propagates only the first order moment instead of the full multi target posterior which makes it a computationally cheaper alternative to optimal multi-target filtering The PHD filter only gives the estimates of the states of targets that are present in the scene at any time step. It keeps no records of the target identities and hence does not produce tracks of individual targets over time. Therefore, certain data association function needs to be incorporated with the PHD filter to produce the desired tracks**.** The method was applied to a video sequence containing multiple faces (QCIF format).

1-3 Features of target model

Object trackers can be guided by a variety of features and selecting the right ones plays a critical role in this context. The uniqueness of a feature is a key property so that the objects can be easily distinguished in the feature space [12]. Some additional discussions and related work about these types of features are presented next [6].

1-The color features

One advantage of the color features is that they are robust against appearance changes and complex backgrounds.

2- The local features

Local discontinuity patches defining points of interest form a common type of local features, such as the ones found in the KLT methods.

3- Other features and hybrid schemes

Some authors, such as Lee et al [13]. consider an appearance (image-based) face tracker as opposed to a feature based tracker. Their appearance model is a piece wise linear approximation of the appearance manifold, i.e., a collection of affine subspaces in the image space.

Figure 1.1: Shows face detection.

1-4 Human tracking using a single camera

A single-camera human tracking task used in the fields of computer vision and sensory biology to track the movement of people in software or high-motion environments. There are different ways to achieve this task, depending on the type of camera available. An example of using a single camera is MOTChallenge [14]: a benchmark for single-camera Multiple Object Tracking (MOT) launched in late 2014, to collect existing and new data and create a framework for the Standardized evaluation of multiple object tracking methods. The benchmark is focused on multiple people tracking, since Pedestrians are by far the most studied object in the tracking community, with applications ranging from robot navigation to self-driving cars.

In the context of single-camera tracking, which typically involves a relatively straightforward scenario of tracking a single object of interest within a camera's field of view, the tracking process can be simplified. It entails searching for the desired object within the expected region of the scene and establishing a coherent trajectory by connecting the object's positions over time. By employing this approach, the system can effectively track and monitor the movement of the targeted object within the captured video footage [15]. Despite the low cost and ease of installation, the problem with single-camera monitoring is the limited visual field [16]. it has been observed that single camera-based systems may struggle to accurately track multiple objects due to horizon limitations, object occlusion, and ambiguous appearances [15].

1-5 Advantages of face detection and tracking

There are many benefits to be gained from combining face detection and tracking. The following applications highlight some of these benefits.

1. Security: It's commonly used in surveillance systems, allowing for the identification and tracking of individuals in real-time, enhancing security measures.
2. Marketing and Analytics: Businesses can use face detection and tracking for demographic analysis and targeted advertising, providing valuable insights into customer behavior.
3. Healthcare: It can assist in medical applications such as patient monitoring or diagnosing certain conditions by analyzing facial expressions.
4. Automated Systems: In robotics and automation, face detection and tracking enable machines to interact with humans more naturally, facilitating applications like social robots or autonomous vehicles.

Overall, face detection and tracking technologies offer a wide range of applications across various industries, enhancing security, convenience, and efficiency.

1-6 Disadvantages of face detection and tracking

While face detection and tracking offer numerous benefits, they also pose several potential disadvantages:

1. Privacy Concerns: There are significant privacy implications associated with face detection and tracking, as it involves the collection and analysis of individuals' facial data without their consent, raising concerns about surveillance and potential misuse of personal information.
2. Security Risks: Face detection and tracking systems can be vulnerable to hacking or spoofing, where malicious actors can manipulate or deceive the system to gain unauthorized access or commit fraudulent activities.
3. Inaccuracy and Error Rates: Despite advancements, face detection and tracking systems are not always accurate, leading to false positives or negatives, which can result in inconvenience or harm, especially in security-critical applications.
4. Intrusiveness: Some people find the idea of constant face detection and tracking intrusive or unsettling, leading to discomfort or resistance to its implementation in public spaces or everyday devices.

Overall, while face detection and tracking technologies offer significant potential, addressing these disadvantages is crucial to ensure responsible and ethical deployment while minimizing risks to individuals' privacy, security, and rights.

1-7 Research objectives

The aim of this research is to develop a system that tracks people and based on their detected faces in videos or camera for various purposes such as security and surveillance. This system is used to identify people and repeatedly track them across consecutive frames.

1. The first objective of the research is to develop a face detection system that accurately and effectively identifies faces in videos or CCTV Camera.
2. The second objective is to track the person(s) based on detected faces by drawing a tracking map.
3. The third objective is to provide a warning when a person enters a potentially dangerous area for enhanced safety and security conditions.

Chapter Two

The suggested project

2-1 Introduction

In this chapter, we will discuss the steps of the project that will focus on tracking people in a scene using a face detection algorithm. Figure 2.1 shows the block diagram of the project. The steps of the project are explained as follows.

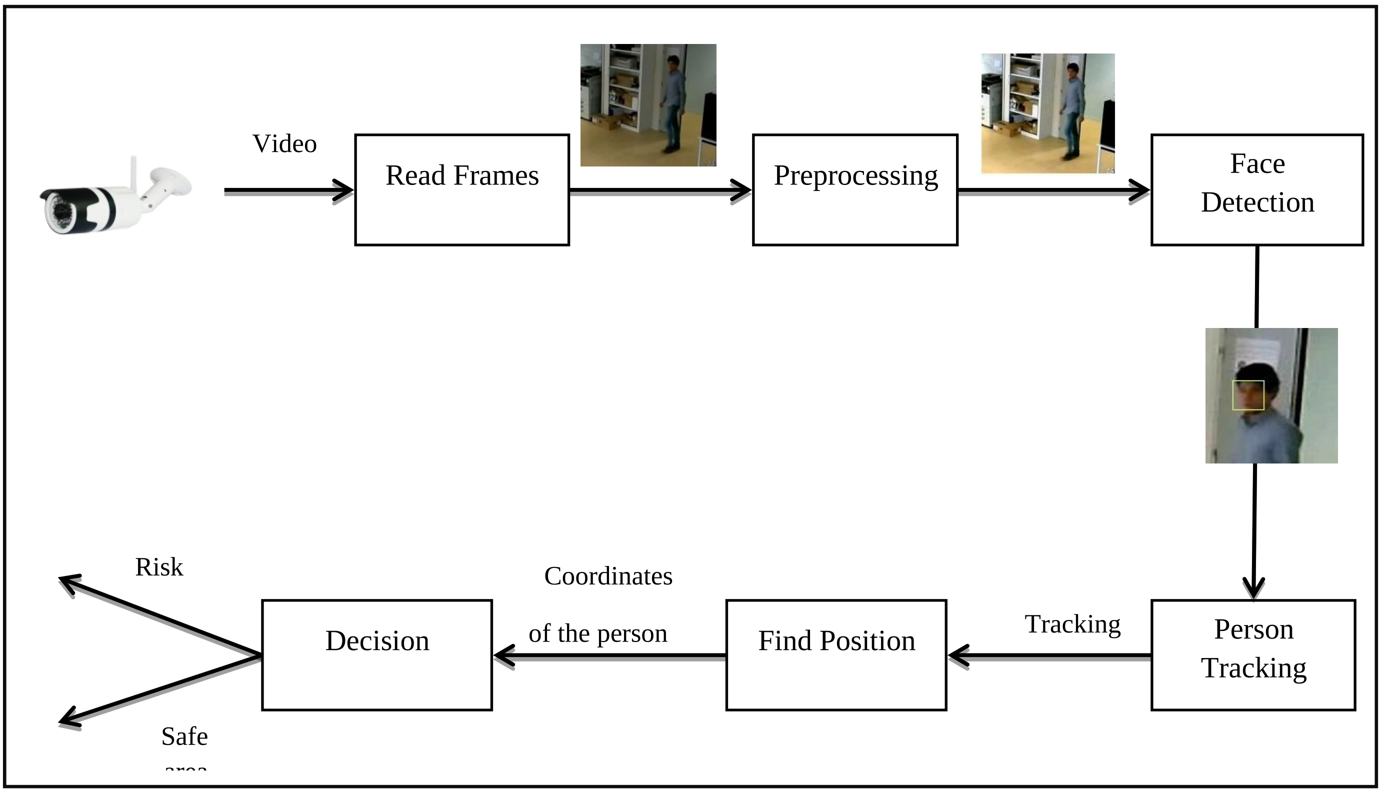

Figure 2.1: The block diagram of the project.

2-2 Video capturing

In the first step, a video camera is used to capture a video sequence. This video source will serve as the input for the face detection algorithm. It is crucial to ensure that the captured video is stable, clear, and free from noise to improve the accuracy of the face detection algorithm. Proper lighting conditions and camera settings can also play a crucial role in capturing high-quality video footage for accurate person detection.

Figure 2.2 shows, aclosed circuit television (CCTV) camera that can be used for video capturing.

Figure 2.2: Closed Circuit Television (CCTV).

2-3 Read frames

After the frame capture process, we read the frames from the video or camera. This procedure is used to extract each frame from the video and handle them individually, allowing each frame to be analyzed and processed separately. It is an important and essential step in the process of detecting faces/people in the video.

2-4 Preprocessing

This step aims to process each frame of the video before detection, in order to improve the lighting accuracy and video quality, enabling the detector to detect a larger amount of objects.

Several functions were used in this context, including controlling the video size, as the size of the video varies and can be very large, and the complexities will vary from one video to another, so it is preferable to set a fixed size such as (600 x 1000). Furthermore, a function for adjusting frame contrast was employed to enhance the differences between dark and light areas in the image or video, potentially improving facial feature visibility and increasing face detection accuracy. Additionally, a noise reduction function was also utilized because frames of images or videos sometimes suffer from noise, and this function is used to estimate and remove noise from the image. Figure 2.3 shows the results of using image enhancement techniques.

(a) original image

(b) using imadjust

(c) using wiener2

Figure 2.3: The results after optimization, where (a) Image before enhancement (original) , (b) Image after optimization using(imadjust) and (c) Image after optimization using(wiener2).

2-5 Face detection

In this step, a process that allows identifying and recognizing faces in video and camera is applied. Advanced techniques are used to detect faces and infer their locations in each frame of the video. This step contributes to distinguishing and identifying the individuals present and following them over time . One of the most common algorithms, i.e. Viola-Jones algorithm is used to detect faces in a video frame. Figure2.4 shows the results of using Viola Jones algorithm for face detection.

(a)Before Face Detection

(b)After Face Detection

Figure 2.4: Shows two pictures: (a) before face detection and (b) after face detection.

As shown in Figure 2.4, faces are successfully detected. Successful detection under less than ideal conditions demonstrates the system's efficiency and accuracy in image analysis. This enhances the effectiveness of facial recognition applications and demonstrates advances in person recognition technologies. However, there may be some detection errors due to lighting and facial tilt that can lead to loss of recognition of human facial features.

2-6 Person tracking

In this step, we monitor and track the movement of people in the video or image across multiple frames using the KLT algorithm, which is a common algorithm in object tracking applications. People tracking aims to locate people, track their movement, and determine their paths over time, whether based on general movement or on special features such as color or pattern. People tracking is used in a variety of applications such as security monitoring, human behavior analysis, smart surveillance, and many other uses that require identifying and monitoring the movement of people.

2-7 Find the position

This step refers to the process that aims to search for the locations of the central points of objects detected by the face detector. The points that can be found can be in the middle of faces, such as the middle of the eyes or the middle of the face itself. The following mathematical equations were used to calculate the midpoints.

$$x = x_1 + (x_1 + x_2) \div 2 \qquad (1)$$

$$y = y_1 + (y_1 + y_2) \div 2 \qquad (2)$$

where x_1, x_2, y_1, y_2 are the coordinates of the face bounding box.

Then, the new calculated coordinates of the center point were converted into real points using the following equations:

$$z = (a_1 - a_2) \div c \qquad (1)$$

$$m = (b_1 - b_2) \div r \qquad (2)$$

where a_1, a_2, b_1, b_2 are the maximum and minimum dimension of the frame width and length respectively, and r, c are number of rows and columns in the original frame.

2-8 Detecting the urgent location

In the final step, a map is drawn based on the previously calculated centroids. After that, the alarm system was organized so that it issues an immediate alert if someone enters the designated danger zone. In this case, the frame was divided into specific risk and safe areas, which are the lower area and the upper area. Rapid and appropriate action can be taken in emergency situations to ensure the safety of individuals in those vulnerable areas. Figure 2.5 shows the (marked) areas that have been enforced.

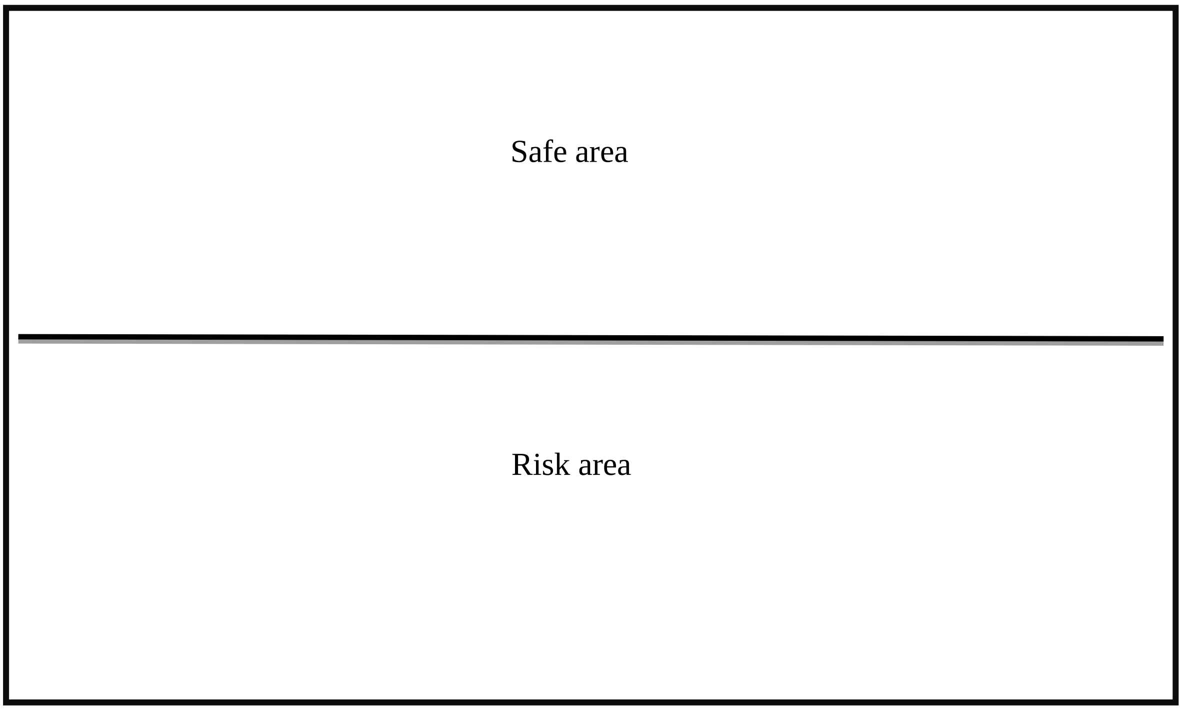

Figure 2.5: The hypothesized risk areas.

Chapter Three

Results and Discussion

3-1 Introduction

In this chapter, we will conduct several experiments using the dataset to detect the faces and track the person. Different scenarios, such as an environment with different lighting or different angles, are suggested to verify the performance of the project. These scenarios are also used to evaluate the system's ability and efficiency to recognize different faces in each trial.

3-2 Experiment setup

The proposed project is run on LENOVO DESKTOP - IVTOF09 with Intel processor Core i5-3320M 2.60GHz and 4.00 GB RAM. This project is written using MATLAB 2021 installed on Windows 64-bit operating system.

3-3 Dataset

The experiments of the proposed project is conducted on WiseNET dataset [17]. The WiseNET dataset provides multi-camera multi-space video sets, along with manual and automatic people detection/tracking annotations and the complete contextual information of the environment where the network was deployed. The WiseNET dataset consists of reassembling 11 video sets, comprising 62 individual videos, recorded using 6 internal cameras deployed in multiple spaces. The video sets represent over an hour of video footage, including tracks for 69 different individuals, capturing various human motions such as walking, standing/sitting, stillness, entering/exiting a space, and group merging/splitting.

3-4 Single face detection

In this experiment, we will conducted two scenarios, front face and side face detection.

In the first scenario, the camera is placed in front of the subject in a well-lit environment, providing enough light to capture the complex and subtle facial details of an individual's face, such as fine lines, wrinkles, skin texture, and subtle differences in skin tone and expressions. Despite encountering light conditions, such asdark background, the experiment effectively identified and track the facial movements. The complexities involved in controlling dynamic facial features and their variations highlighted the challenges encountered, but yielded accurate and satisfactory results. Finally, the project deliverables also include a warning message to let us know that the person is in a safe or unsafe area. The results of this experiment are shown in Figure 3.1.

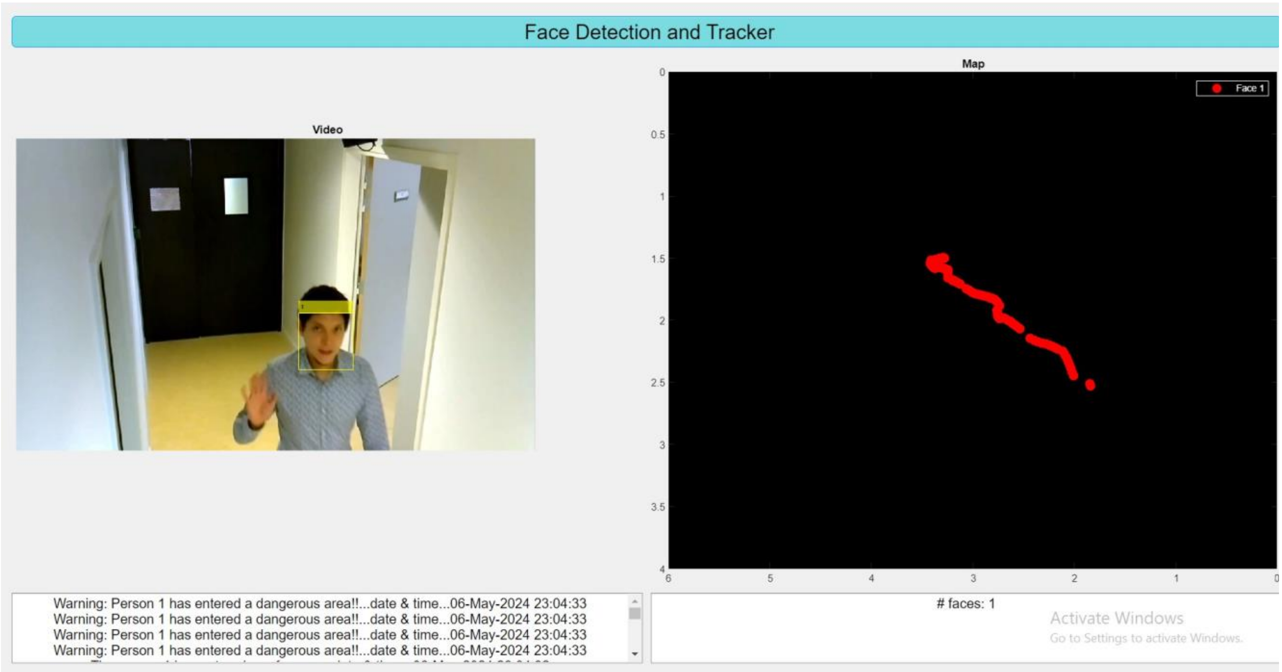

Figure 3.1 The results of single face detection(first scenario).

In the second scenario, a side face detection and tracking is implemented. The results of this experiment are shown in Figure 3.2. The results of this experiment reveal that the side face detection process presented greater difficulties compared to the first scenario. which presents challenges in both the face detection and tracking processes. Despite obstacles such as poor lighting and missing facial details , the system successfully overcomed these conditions and achieved accurate and noteworthy results. This experiment demonstrate the system's efficiency in dealing with diverse conditions and ensures accurate and continuous detection and tracking.

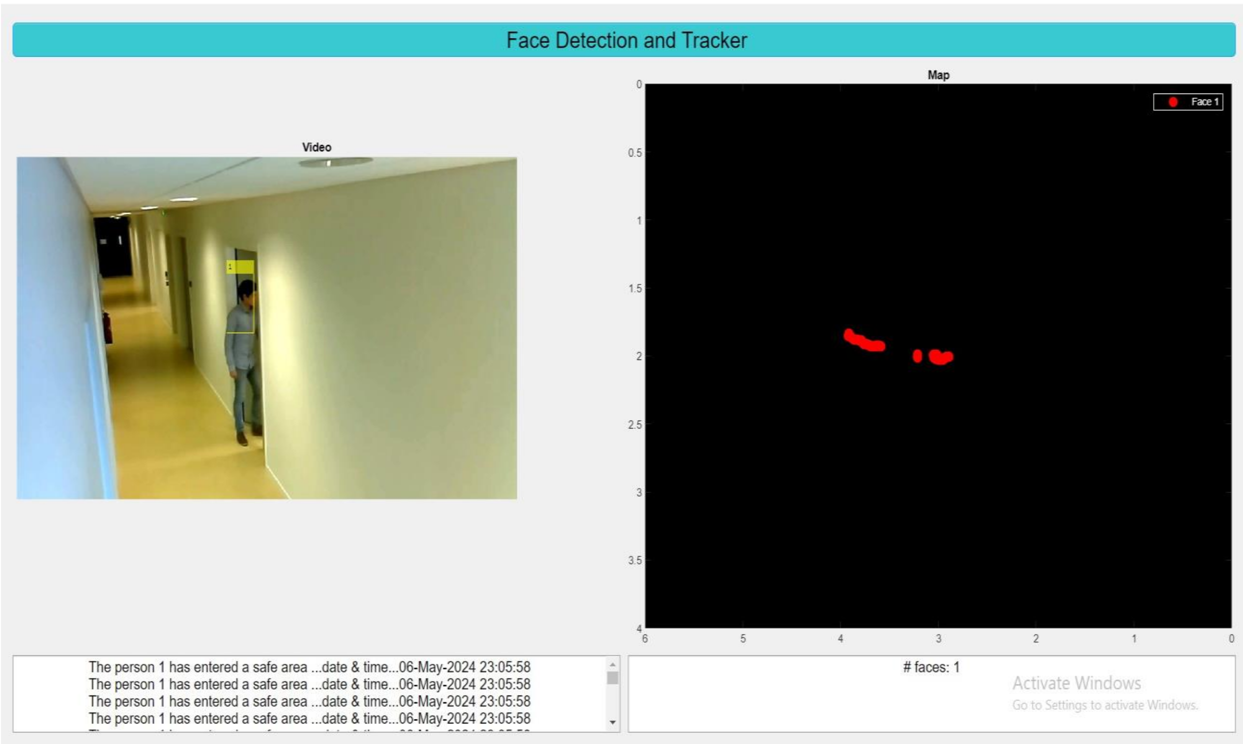

Figure 3.2: The results of single face detection(second scenario).

3-5 Detect multiple faces

In the second experiment we will also run two scenarios.

In this first scenario, this experiment was used to verify the method's ability to address the challenges of light conditions and multiple side face detection. Despite these challenges and others in the scene, the system was able to successfully detect and track multiple faces.

The experiment began with changes in lighting and camera angles, and identifying more than one face, which increases the complexity of the task. However, the system was able to successfully detect and track large portions of faces. This reflects the system's ability to adapt to difficult circumstances and deal with many challenges.

In short, the second experiment represents a remarkable achievement in the field of face detection and tracking, despite the challenges that can be found in the scene. The results shown in Figure 3.2 demonstrate the system's ability to handle volatile environments and achieve excellent performance.

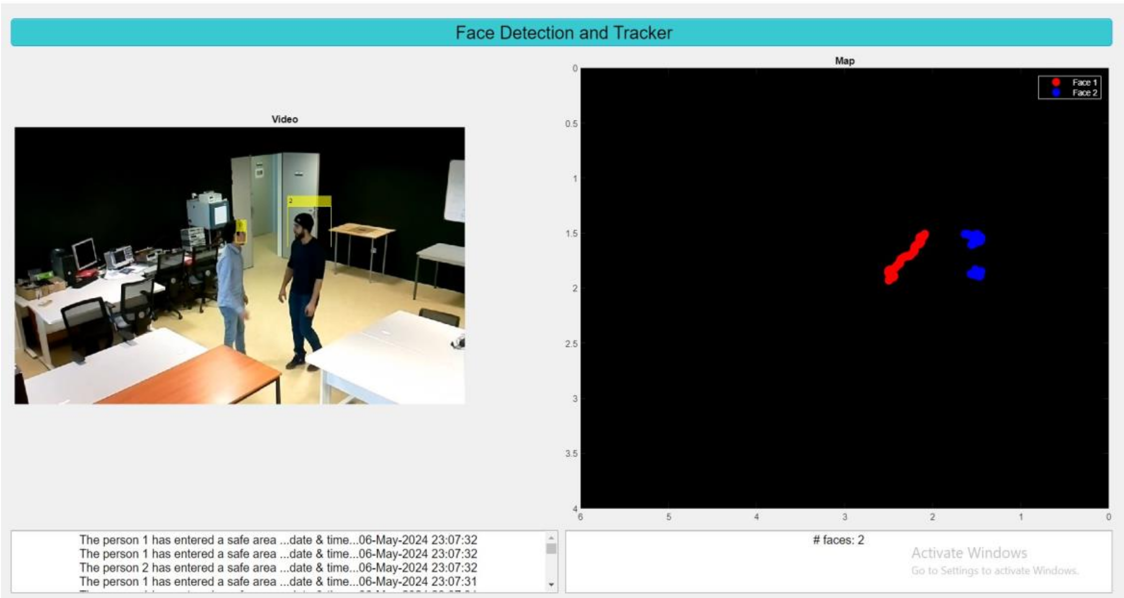

Figure 3.3 The results of multiple face detection(first scenario).

In the second scenario, the experiment is also used to detect multiple faces rather than one. This scenario of the experiment is similar to the previous scenario except that the light conditions are considered good compared to the first experiment. However, the project succeeded in detecting the faces. Although various challenges were encountered, including side face detection, the results are documented and represented visually in Figure 3.4.

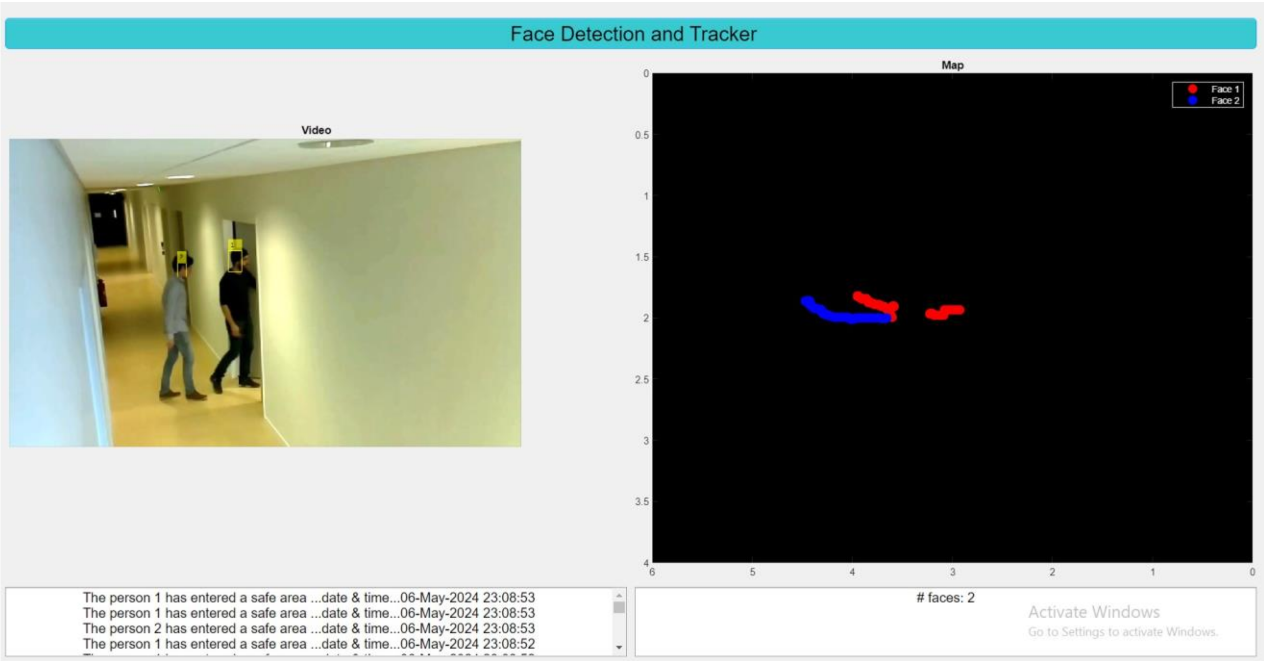

Figure 3.4: The results of multiple face detection(second scenario).

From the set of the conducted experiments, we found that the method was able to detect faces despite the challenges present in the scenes. The project deal with the challenges of light conditions, angle of faces, field of view, and the distance between the face and the camera. However, the project overcome all these challenges and show good results.

Chapter Four

Conclusions and Future work

4-1 Conclusions

Detection and tracking operations were successfully achieved despite numerous challenges, of which lighting effects and the difficulty of detecting faces from the side were among the most notable. Using these technologies, we can effectively identify high-risk areas. The conducted experiments demonstrate the effectiveness of the camera used and the efficiency of the detector in carrying out its tasks accurately and effectively. However, we should continue to work on improving these techniques to reduce the effects of external factors, increase their accuracy, and increase their adoption in various scenarios.

4-2 Future work

One of the challenges we face in this project is the field of view covered by a single camera, as it does not provide all the required information about people's movements. In addition, there are some detection errors. A possible solution to this issue involves using more than one camera.

Integrating an online camera system can greatly enhance monitoring capabilities by enabling real-time tracking and analysis of individuals in high-risk areas. The continuous video feed from an online camera allows for immediate identification of faces and movements, enhancing overall surveillance measures. Furthermore, the proposal to implement deep learning algorithms, such as YOLO (You Only Look Once), can significantly boost the accuracy and efficiency of the face detection and people tracking system. YOLO's speed and precision in object detection tasks make it a fitting choice to improve the system's performance in identifying and tracking individuals effectively.

References

[1] Mamieva, D., Abdusalomov, A.B., Mukhiddinov, M. and Whangbo, T.K., 2023. Improved face detection method via learning small faces on hard images based on a deep learning approach. Sensors, 23(1), p.502.

[2] Katke, K.S., 2020. Face detection using Viola Jones algorithm and MLP technique. Int. J, 5, pp.2456-0774.

[3] Tseng, S., 2003. Comparison of holistic and feature based approaches to face recognition. Masters of Applied Science in Information Technology thesis, School of Computer Science and Information Technology, Faculty of Applied Science, Royal Melbourne Institute of Technology University, Melbourne, Victoria, Australia.

[4] Garg, D., Goel, P., Pandya, S., Ganatra, A. and Kotecha, K., 2018, November. A deep learning approach for face detection using YOLO. In 2018 IEEE Punecon (pp. 1-4). IEEE.

[5] Rosiani, U.D., Choirina, P. and Shoumi, M.N., 2021, February. Micro-expression recognition based on motion detection method. In IOP Conference Series: Materials Science and Engineering (Vol. 1073, No. 1, p. 012069). IOP Publishing.

[6] Zhang, T. and Gomes, H.M., 2014, March. Technology survey on video face tracking. In Imaging and Multimedia Analytics in a Web and Mobile World 2014 (Vol. 9027, pp. 87-98). SPIE.

[7] Comaniciu, D., Ramesh, V., Meer, P., "Real-time tracking of non-rigid objects using mean shift", IEEE Conference on Computer Vision and Pattern Recognition, pp.142-149 (2000).

[8] Bradski, G. R., "Computer vision face tracking for use in a perceptual user interface," Intel Technology Journal, 2nd Quarter (1998).

[9] Allen, J. G., Xu, R., Jin, J. S., "Object tracking using CamShift algorithm and multiple quantized feature spaces," Proceedings of Pan-Sydney Area Workshop on Visual Information Processing (VIP 2003).

[10] Comaniciu, D., Ramesh, V., Meer, P., "Kernel-based object tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5 (2003).

[11] Kalal, Z., Mikolajczyk, K., Matas, J., "Tracking-learning-detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.6, No.1 (2010).

[12] Yilmaz, A., Javed, O., Shah, M., "Object tracking: A survey," Journal of ACM Computing Surveys, Vol. 38, Issue 4, Article No. 13 (2006).

[13] Lee, K. C., Ho, J., Yang, M. H., Kriegman, D., "Visual tracking and recognition using probabilistic appearance manifolds," Journal of Computer Vision and Image Understanding, Vol. 99, Issue 3, pp. 303-331 (2005).

[14] Dendorfer, P., Osep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S. and Leal-Taixé, L., 2021. Motchallenge: A benchmark for single-camera multiple target tracking. International Journal of Computer Vision, 129, pp.845-881.

[15] Amosa, T.I., Sebastian, P., Izhar, L.I., Ibrahim, O., Ayinla, L.S., Bahashwan, A.A., Bala, A. and Samaila, Y.A., 2023. Multi-camera multi-object tracking: a review of current trends and future advances. Neurocomputing, 552, p.126558.

[16] Kawasaki, A., Hung, D.H. and Saito, H., 2014, August. Human trajectory tracking using a single omnidirectional camera. In Proceedings of the 16th Irish Machine Vision and Image Processing (IMVIP) Conference (pp. 157-162).

[17] Marroquin, Roberto, Julien Dubois, and Christophe Nicolle. "Ontology for a Panoptes building: Exploiting contextual information and a smart camera network." Semantic Web 9.6 (2018): 803-828.