JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CCTV-Powered Facial Authentication for Visitors

1'st Ms. Mathumitha M
Dept. Information Technology (IT)
M.A.M College of Engineering and Technology (MAMCET)
(Anna University of Affiliation)
Trichy, India
mathumitha.it@mamcet.com

2'nd Afza fathima S
Dept. Information Technology (IT)
M.A.M College of Engineering and Technology (MAMCET)
(Anna University of Affiliation)
Trichy, India
afzafathima.it20@mamcet.com

3'rd Ishwarya K
Dept. Information Technology (IT)
M.A.M College of Engineering and Technology (MAMCET)
(Anna University of Affiliation)
Trichy, India
ishwarya.it20@mamcet.com

4'th Keerthika E
Dept. Information Technology (IT)
M.A.M College of Engineering and Technology (MAMCET)
(Anna University of Affiliation)
Trichy, India
keerthika.it20@mamcet.com

Abstract*—Recent advancements in image recognition technology, particularly deep learning, have revolutionized security and home services by incorporating biometric data like fingerprints, iris scans, and face recognition. Among these, face recognition-based user authentication methods have gained significant traction. This study introduces a visitor authentication system using CCTV, Jetson Nano, and a webcam. In the preprocessing phase, CCTV captures face data with 7 key features for person identification. The collected dataset undergoes annotation and deep learning-based facial feature detection. If four or more features are detected, the data is classified as a person, and facial details are matched with stored user data using 81 feature vectors. Furthermore, security enhancements include logging visitor faces, visitor count, and visit times, bolstering the access control system. The paper implements this system on a Jetson Nano and evaluates its performance, focusing on accuracy and detection speed. The tiny-YOLOv3 model on Jetson Nano proves effective, achieving real-time face authentication with an average detection speed of 6.5 FPS and 86.3% accuracy. This study showcases a deep learning-based system for visitor authentication and access control, demonstrating its capabilities in real-time verification and user management.

***Keywords*— Face recognition, Deep learning, Visitor authentication, Jetson nano**

I. INTRODUCTION

Closed-circuit television (CCTV) systems have emerged as pivotal tools in modern security infrastructure, offering unparalleled surveillance capabilities and real-time monitoring across various environments. With advancements in image recognition technology, particularly through the integration of deep learning algorithms, CCTV systems have evolved beyond mere video recording devices to sophisticated platforms capable of biometric data analysis and user authentication. The convergence of CCTV technology with biometric information such as fingerprints, iris scans, and face recognition has revolutionized security protocols, enabling robust access control systems and enhancing the overall safety of both public and private spaces. In this context, the utilization of CCTV in conjunction with advanced hardware components like the Jetson Nano and webcams represents a significant stride towards creating intelligent surveillance solutions that not only capture and store video footage but also perform real-time analysis and authentication tasks. This paper delves into the development and evaluation of a visitor authentication system leveraging CCTV, Jetson Nano, and deep learning algorithms, showcasing the transformative potential of integrating cutting-edge technologies in the realm of security and access control. CCTV plays a crucial role in visitor authentication by serving as the primary source of visual data for identifying and verifying individuals accessing a premise or facility. The integration of CCTV with advanced technologies such as deep learning algorithms enhances its capabilities in accurately recognizing and authenticating visitors based on their facial features. CCTV captures high-resolution video footage of visitors entering and interacting within the monitored area. This footage serves as the foundation for the authentication process, providing visual data that can be analysed and processed to extract relevant biometric information. CCTV captures high-resolution video footage of visitors entering and interacting within the monitored area. This footage serves as the foundation for the authentication process, providing visual data that can be analysed and processed to extract relevant biometric information. CCTV enables real-time monitoring and verification of visitors as they move through different areas or checkpoints within a premise. This continuous monitoring ensures that only authorized individuals are granted access, enhancing security and preventing unauthorized entry. Moreover, CCTV plays a vital role in recording and storing visual evidence of visitor authentication events. This data logging feature enables security personnel to review and audit access logs, track visitor movements, and investigate security incidents if necessary. Overall, the role of CCTV in visitor authentication is multifaceted and instrumental in ensuring robust access control measures. By leveraging advanced facial recognition technology and real-time monitoring capabilities, CCTV systems contribute significantly to enhancing the security, efficiency, and reliability of visitor authentication processes in various environments, including commercial buildings, residential complexes, educational institutions, and public facilities. Hence, we propose an advanced system that leverages AI technology integrated with CCTV to effectively identify and manage visitors based on situational analysis.

The proposed system incorporates state-of-the-art techniques such as You Only Look Once (YOLO) and OpenCV, utilizing live image feeds from CCTV cameras positioned at entry points to determine various critical factors. These include the recognition of human faces, identification of potential intruders, and detection of suspicious behaviour such as attempts to steal access credentials by observing registered users. By employing YOLO and OpenCV algorithms, our system can promptly analyse incoming video streams and make accurate assessments regarding visitor identity and intentions. For instance, it can identify instances where an individual is attempting to gain unauthorized access or engage in illicit activities like shoulder surfing to obtain door codes. Upon detecting such anomalies, the system triggers an immediate alarm notification to alert authorized users, providing them with real-time insights into potential security threats. One of the key functionalities of our system is its ability to detect and notify users when an unauthorized person is in close proximity to a registered individual, particularly in scenarios where theft or intrusion is suspected. This proactive approach is crucial in preventing security breaches and safeguarding the privacy and safety of residents or users.

Notably, the system also addresses emerging security concerns, such as incidents where outsiders attempt to gain access to residential or commercial premises, especially during vulnerable times such as nighttime or when occupants are alone. Furthermore, our system facilitates seamless communication and collaboration with building management or security personnel by automatically transmitting pertinent data, including photographs and access logs, in response to potential security incidents. This information not only aids in preventing unauthorized access but also serves as valuable evidence during incident investigations and legal proceedings, ensuring swift and effective responses to security breaches.

REQUIREMENTS THAT THE SYSTEM SHOULD

SATISFY : Technological advancements in agriculture have transformed the way we cultivate crops, manage resources, and sustainably feed the world. From precision farming techniques to artificial intelligence (AI) and robotics, these innovations are revolutionizing every aspect of the agricultural sector.

1. **The system must differentiate between an intruder and a companion:** When the companion is a family member, the alarm notification should specify them as such. However, if the companion is not a family member, a lengthy password is necessary to prevent unauthorized observation.

2. **The system is designed to detect any attempt to observe the password surreptitiously:** It utilizes advanced algorithms to analyze the companion's eye movements and assess whether they are attempting to peek over the registered user's shoulder. If the system detects such behavior, it immediately triggers an alarm prompting the user to enter a longer password for enhanced security measures. In cases where a short password is entered, the system generates a notification message alerting the user that the password length is insufficient, thereby prompting them to choose a longer and more secure password.

3. **The system is designed to authenticate the faces of registered individuals, specifically family members whose photos are pre-registered:** Using OpenCV for facial recognition, the system verifies the identity of visitors. Upon successful authentication, the door can be unlocked using multi-factor authentication, requiring the input of a password at the front door for added security.

4. **The system must encompass a range of essential services:** remote door opening and intruder detection. Additionally, it should have the capability to notify the system manager via smartphone upon the family's return home, offer real-time access information checks, analyze access patterns, and provide supplementary services in cases of prolonged absence.

This study employs cutting-edge technology, utilizing a Jetson Nano along with YOLO, a deep learning model specializing in object detection and real-time visitor recognition. Through this integration, we aim to develop a comprehensive system capable of identifying intruders, validating authorized visitors, and detecting suspicious activities such as shoulder surfing. Apart from the Jetson Nano, which is commonly used in user authentication based on face recognition using CCTV technology with deep learning algorithms like multi CNN(Convolutional Neural Network), there are several other microprocessors that can be utilized for similar applications. These microprocessors provide varying levels of performance, power efficiency, and integration capabilities suitable for different use cases in user authentication and facial recognition systems. One popular choice is the Raspberry Pi, known for its affordability, versatility, and community support, making it ideal for DIY projects and prototypes in smaller-scale deployments. The NVIDIA Jetson Xavier NX stands out for its high computational power and suitability for real-time facial recognition tasks, making it a preferred option for security systems requiring rapid processing and accuracy. The Intel Movidius Neural Compute Stick offers hardware acceleration for deep learning models, enhancing facial recognition capabilities in edge computing environments. Similarly, the Google Coral Dev Board, featuring the Edge TPU (Tensor Processing Unit), is well-suited for deploying deep learning models for user authentication and facial recognition in edge computing setups. Additionally, Arduino boards, while not as powerful for heavy-duty deep learning tasks, can still be integrated into simpler facial recognition systems for basic authentication needs or as part of a larger system combining multiple processors for optimized performance.

This paper makes several significant contributions to the field:

1. It presents a cost-effective system developed as a standalone solution, eliminating the need for server-based infrastructure and reducing overall expenses.

2. Through experimentation with both Raspberry Pi and Jetson Nano within the same environment, the paper highlights variations in accuracy, offering valuable insights into the performance of different microprocessors in user authentication systems.

3. The paper successfully achieves the objective of visitor authentication by implementing robust face detection and face recognition mechanisms.

4. Furthermore, the system is validated to be secure against various potential attacks, ensuring the safety of user data, providing convenience in authentication processes, and offering a seamless visitor authentication experience.

PRIOR STUDIES:

A. USER AUTHENTICATION: User authentication based on face recognition using CCTV technology is a cutting-edge project that aims to enhance security and convenience in access control systems. By leveraging deep learning algorithms and advanced image recognition techniques, the system can accurately identify and authenticate users based on their facial features captured by CCTV cameras. This eliminates the need for traditional authentication methods like passwords or access cards, reducing the risk of unauthorized access and enhancing overall security. Additionally, the use of CCTV technology ensures real-time monitoring and verification, making the authentication process efficient and seamless for users. Overall, this project represents a significant advancement in user authentication systems, offering a secure, reliable, and user-friendly approach to access control.

1. KNOWLEDGE BASED AUTHENTICATION:

Knowledge-Based Authentication (KBA), also referred to as knowledge questions authentication, is an identity verification method that relies on a series of questions about personal information to ensure that the individual accessing a place or an account is the legitimate owner and not an unauthorized user. The essence of KBA lies in the fact that only the true account holder would possess the knowledge required to answer these questions accurately. KBA authentication can be categorized into two main types: static and dynamic. Static KBA involves asking predetermined questions about personal details or historical information, while dynamic KBA adapts the questions based on the user's behavior or recent activities, such as recent transactions or account interactions. Despite its historical usage, particularly in password recovery processes, KBA has inherent vulnerabilities and often leads to high friction for users. As cyber threats evolve, relying solely on knowledge-based authentication may not provide sufficient security against sophisticated attacks, prompting organizations to explore more robust authentication methods, such as multi-factor authentication (MFA), to enhance account security and mitigate risks associated with KBA vulnerabilities. he vulnerabilities associated with the MD5 hash function, known for generating 128-bit hash values, have necessitated its widespread replacement by more secure hash functions like SHA-256 and SHA-512. Despite these advancements, password security remains vulnerable to various threats. For instance, passwords can still be compromised through shoulder surfing or recording inputs, especially with technologies like Google Glass. Social engineering attacks targeting user information or brute force attacks on passwords are also persistent threats. Furthermore, in the event of a data breach where password hashes are exposed, Rainbow Table attacks can be conducted to retrieve user passwords from pre-generated hash values. These security risks highlight the ongoing challenges in maintaining robust password security measures.

2. TOKEN AUTHENTICATION: Token authentication is a popular method used to verify the identity of users accessing systems or services securely over networks. It operates on the principle of using a unique token, often generated dynamically, to authenticate users instead of relying solely on traditional methods like passwords or biometric data. The working of token authentication involves several key steps. Firstly, when a user attempts to access a system or service, they are required to provide their credentials, which may include a username and password. Upon successful verification of these credentials, the system generates a token specific to that user's session. This token is typically a string of characters or a cryptographic key that serves as proof of the user's identity and authorization to access the system. Next, the token is transmitted securely to the user's device, often through encrypted communication protocols such as HTTPS or SSL/TLS. The token is stored temporarily on the user's device, usually in a secure storage area such as a cookie or local storage. During subsequent interactions with the system or service, the user presents the token instead of re-entering their credentials. The system validates the token to ensure its authenticity and matches it with the corresponding user session information stored in the server's database. If the token is valid and matches the session data, the user is granted access to the requested resources or functionalities.

Token authentication offers several advantages over traditional authentication methods. Firstly, tokens can have a limited lifespan, known as time-based expiration, or can be invalidated after a single use, enhancing security by reducing the risk of token theft or unauthorized access. Additionally, tokens can be scoped to specific resources or operations, allowing fine-grained control over user permissions and access privileges. Furthermore, token authentication is often used in conjunction with other security measures such as multi-factor authentication (MFA) or biometric authentication to add an extra layer of security. For example, a token generated on a user's mobile device may require biometric verification (e.g., fingerprint or facial recognition) before being accepted by the system. Overall, token authentication is a versatile and effective method for securely verifying user identity and managing access to systems and services, offering enhanced security, flexibility, and usability in modern authentication mechanisms.

3. BIOMETRIC AUTHENTICATION: Amid the challenges posed by the COVID-19 pandemic, there is a growing need for advanced visitor management solutions that can meet the evolving requirements of today's environment. Biometric authentication emerges as a highly effective approach for visitor management providers, offering a range of benefits over traditional methods. Biometric technologies leverage unique biological traits like facial features, voice patterns, or iris scans to verify identities, eliminating the need for physical touchpoints and manual data entry. One of the key advantages of biometric solutions is their mobile compatibility, utilizing the native capabilities of devices such as cameras and microphones for secure and convenient identification. This eliminates the necessity for specialized equipment, streamlining the authentication process for visitors and employees alike. Instead of typing in information on a touchpad, individuals can simply present their face, voice, or iris to gain access or notify personnel of their arrival, enhancing efficiency and minimizing contact points. The adoption of biometrics also significantly enhances the speed of visitor management workflows. By replacing time-consuming manual entry with swift biometric recognition, users can navigate through check-in procedures more efficiently, saving time and improving overall convenience. Moreover, the accelerated process reduces congestion in shared areas like lobbies, contributing to enhanced security and compliance with social distancing guidelines. Beyond addressing health concerns related to physical touchpoints, biometric authentication strengthens building security by eliminating reliance on vulnerable methods like passwords or keycards. These traditional access mechanisms are prone to theft or unauthorized use, whereas biometrics offer a more robust and reliable form of identification based on unique biological characteristics. Integrating biometrics into visitor management systems provides a higher level of security, aligning with modern security standards and offering a superior alternative to conventional authentication methods

B. ADVANCED ARTIFICIAL INTELLIGENCE AND IMAGE PROCESSING TECHNOLOGIES

1. OBJECT DETECTION TECHNOLOGY: The YOLO (You Only Look Once) system is a Convolutional Neural Network (CNN), a type of deep neural network designed specifically for real-time object detection. CNNs operate by processing input images as structured arrays of data, allowing them to identify patterns and features within the images. YOLO stands out for its remarkable speed compared to other networks while still maintaining high accuracy levels. What sets YOLO apart is its ability to analyze the entire image during testing, leveraging the global context to make accurate predictions. In essence, YOLO and similar convolutional neural network algorithms evaluate regions within an image and assign scores based on their resemblance to predefined classes. Regions with high scores are identified as positive detections of specific classes they closely match. For instance, in applications like self-driving cars, YOLO excels at identifying various types of vehicles by analyzing regions within the video that align with predefined vehicle classes. This scoring mechanism, which involves regional proposals, enables YOLO to achieve precise and efficient object detection across diverse scenes, making it a valuable tool in real-time image processing and artificial intelligence applications. Some users may opt to continue using YOLOv3 instead of newer versions like YOLOv4 or YOLOv5 based on specific project requirements. One key factor influencing this decision is the stability and maturity of YOLOv3. Having been available for a longer period, YOLOv3 has undergone thorough testing and validation across various applications, garnering a reputation for stability and reliability. This extensive experience instil confidence, particularly in scenarios where the latest functionalities introduced in newer iterations are not essential. Another consideration is the practical aspect, such as model size and deployment feasibility. YOLOv3 is known for its relatively smaller model size compared to later versions, making it well-suited for deployment on edge devices with limited storage capacity or constrained bandwidth. Users prioritizing compactness and ease of deployment may find YOLOv3 more advantageous over newer versions that may require more storage space and computational resources. Therefore, the decision to stick with YOLOv3 is often driven by a balance of stability, practicality, and compatibility with project-specific requirements.

2. OBJECT DETECTION TECHNOLOGY IN THE REAL - TIME ENVIRONMENT:

In recent years, deep learning has made remarkable strides in advancing object detection technologies. The categorization of detectors into one-stage and two-stage detectors revolves around their utilization of region proposals. Two-stage detectors like R-CNN, Fast R-CNN, and Faster R-CNN rely on region proposals for precise object detection. While these methods demonstrate accurate detection capabilities, they often suffer from larger model sizes and slower detection speeds. On the other hand, one-stage detection methods such as SSD (Single Shot Multi Box Detector) by Liu et al. and the YOLO (You Only Look Once) series directly predict object locations and categories, resulting in faster detection speeds albeit with slightly reduced accuracy. To cater to the demand for reduced computational complexity and model parameter size, lightweight networks and detectors like MobileNetV2-V3, ShuffleNetV1-V2, YOLOv4-tiny, and YOLOv7-tiny have been developed, offering rapid detection capabilities ideal for deployment on embedded hardware devices. The traditional convolutional neural networks (CNNs) may encounter challenges in detecting occluded objects due to their fixed and limited receptive fields. The emergence of Transformers has significantly expanded the capabilities of neural network models by capturing and leveraging long-range dependencies, thereby enhancing overall performance and addressing CNN limitations. Notably, Dosovitskiy et al. introduced the Vision Transformer, employing Transformer blocks as the foundational architecture and showcasing the effectiveness of Transformers in computer vision tasks. Building upon this, Liu et al. proposed Swin-Transformer to tackle diverse scale challenges in visual elements, achieving notable success in handling scale differences in visual data through hierarchical feature integration and advanced technologies. Despite the impressive modelling prowess of Transformers, they may overlook local image details. Therefore, a promising approach involves combining Transformers with CNNs to enhance robust feature extraction. Noteworthy contributions like BoTNet, which leverages self-attention in the final three blocks, and CvT, an improvement upon ViT (Vision Transformer) by integrating convolutional tokens, showcase the potential of hybrid models. However, it's essential to note that the self-attention mechanism in Transformers can lead to increased computational load and memory usage, posing challenges for deployment in resource-constrained environments like classroom camera systems.

3. END-TO-END FACE RECOGNITION SYSTEM FOR BIOMETRIC AUTHENTICATION:

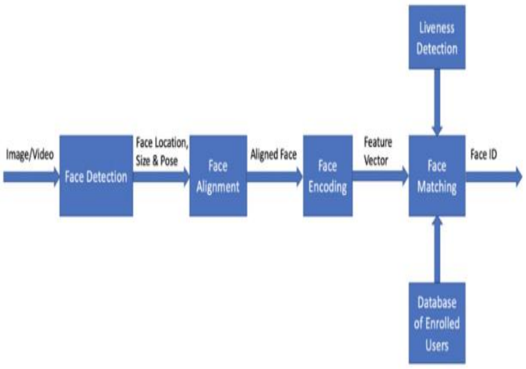
Face recognition technology has become a ubiquitous part of our daily lives, especially with the widespread use of smartphones. Many modern smartphones utilize face recognition technology as a secure method to unlock devices, providing a robust layer of protection for personal data. This technology is instrumental in safeguarding sensitive information, ensuring that even if a phone is stolen, unauthorized access to data remains unattainable. Beyond device security, face recognition is finding applications in a broad range of domains such as safety, security, and payments, showcasing its versatility and importance in enhancing various aspects of our lives. The Global Artificial Intelligence Accelerator (GAIA) team at Ericsson has been actively working on a Proof of Concept project focused on enhancing authentication security. Unlike other companies in the market offering commercial face recognition solutions, GAIA predominantly utilized open-source tools to create an AI-driven solution adaptable for mobile and edge devices. Given the constraints such as limited storage and memory on these devices, striking a balance between model complexity, performance, and response time becomes crucial in selecting the most suitable AI models. Another significant consideration is the trustworthiness of AI models used in face recognition. To address this, Ericsson has developed guidelines for trustworthy AI development, underscoring their commitment to ensuring reliability and integrity in their initiatives. The diagram depicted in Figure 1 illustrates the architectural blueprint of an end-to-end face recognition system designed for biometric authentication purposes. This system operates by taking a limited number of images or video frames as input, detecting human faces within them, and subsequently assessing whether these faces correspond to any of the face images stored in the database of enrolled users. Upon a successful match, the person undergoes biometric verification, whereas failure to match results in non-verification. The system comprises four fundamental modules: face detection, face alignment, face encoding, and face matching. Additionally, an optional face liveness check module is incorporated into the pipeline to verify that the authenticated individual is indeed a live person, thus preventing the system from being deceived by a mere photograph of the targeted person. Face detection serves as the initial phase in the processing pipeline, dedicated to locating a face within an image. Unlike identity determination, this step purely focuses on identifying the presence of a face. The Ultralight detector is chosen as the default face detection model due to its exceptional performance in detecting faces from various angles, not limited to just front-facing positions. Additionally, this detector is characterized by its lightweight nature, occupying less than 2MB of storage space, and boasts a rapid face detection speed, with an average inference time of 50 +/- 6 ms on a MacBook Pro 2.6GHz Intel Core i7 equipped with 32GB DDR4 RAM. Face alignment is the next step, after a face is detected in an image. Quite often when a person takes a picture, he or she may not be facing directly towards the camera. However, face alignment can deal with the problem. Even if a face is turned in different directions, the system is still able to tell if it is the same person. More specifically, an algorithm called “face landmark estimation” is applied to locate facial landmarks, i.e., the specific points that exist on every face, such as top of chin, outside edge of each eye, inner edge of each eyebrow, etc. The third step is face encoding. This process identifies key parts of a face through the “eyes of a computer.” As computers can only recognize numbers, a reliable way of converting face images to numbers/measurements was needed to represent each face. Finding a good method of face encoding was a challenging task. Quite often deep learning models, such as the “Convolutional Neural Network (CNN)” model, are trained by using a large database of face images to calculate the best face representation of each face. The goal of this training is to generate nearly the same encodings when looking at two different pictures of the same person, whilst generating quite different measurements when looking at pictures of different people. After exploring many different models, a pre-trained Resnet model provided in Dlib was chosen for the face encoding model of the pipeline. This model was essentially a ResNet-34 model, which was modified by dropping some layers and re-building with 29 convolution layers. This Resnet model takes an image inputs with size 150 x 150 x 3 and represents/encodes each face image as 128-dim measurements. Once the model network was designed, the pretrained model was trained on a dataset of about 3 million faces. The face dataset was mainly derived from the two open-source face databases, the face scrub dataset and the VGG dataset.

Figure 1. Architecture of end-to-end face recognition system for biometrical authentication

C. ACCESS CONTROL SYSTEM: Access control systems, a vital aspect of physical security, are deployed in various locations to prevent unauthorized access, such as in buildings and offices. An illustrative example of this is observed in the management of vehicle entry and exit at parking lots. Here, either a camera captures the car's license plate number for verification against a database or an RFID-based smart card system is utilized. The RFID card, affixed to the car's windshield, facilitates access validation based on registration. This contactless smart card technology extends its convenience indoors, allowing for seamless access within the office premises. Robust security measures, including encryption algorithms, special keys, and mutual authentication, safeguard against unauthorized copying or hacking attempts, particularly in RFID-based systems. Traditional access methods like password-based smart door locks in residential settings have evolved with advancements in biometric authentication technologies. These innovations, such as fingerprint or iris recognition using cameras, verify users' identities against registered profiles, enhancing access control efficiency and security. Moreover, recent developments have focused on incorporating additional authentication methods, such as voice recognition, PIR proximity sensors, or question-answer protocols, for further authentication layers. Deep learning-based face recognition technology has gained prominence in access management systems, enabling swift and accurate user authentication. This technology is gradually replacing conventional access control methods reliant on security personnel or visitor tags, which can be prone to issues like staff shortages or tag misplacement. Particularly noteworthy is the increasing utilization of face recognition amid the COVID-19 pandemic, aiding in tasks like access list creation, mask detection, and temperature measurement for enhanced safety protocols. Furthermore, deep learning-based object detection technology finds application in bolstering security management systems, offering advanced capabilities for threat detection and prevention. For instance, in front door security, face recognition serves as a viable alternative to mitigate password peeping by unauthorized individuals. Integrating face recognition with alarm systems based on proximity detection further enhances security by alerting to the presence of unauthorized individuals attempting access.

II. EXISYTING SYSTEM

The system's architecture can be broadly categorized into registration and processing components, comprising deep learning model training, visitor face detection, face recognition and access verification, and monitoring functionalities. To commence, the deep learning model training segment involves utilizing Tiny-YOLOv3 to train the model using labelled face features datasets encompassing seven classes. In the visitor face detection module, the foundation for the face recognition security system is laid. This segment incorporates a webcam for CCTV functionality, a servo motor for environmental monitoring, and an infrared sensor for human body detection, all interconnected with the Jetson Nano board. Initially, real-time images from the webcam are analysed using the trained model to detect human faces. Upon confirmation of a human face, intricate facial features are identified using the landmark algorithm within the recognition and security segment. Subsequently, the identified facial features are cross-referenced with stored face data in the monitoring system's database. Access is granted if the detected face belongs to a registered member, while unauthorized access triggers an alarm notification to alert the user. Additionally, comprehensive log information including facial data and headcount is wirelessly transmitted and stored in the database via WLAN for future reference and analysis. This seamless integration of registration, detection, recognition, and monitoring processes ensures robust security and efficient management of access control within the system. The system's architecture is designed to scale seamlessly, accommodating future enhancements such as advanced analytics for visitor behaviour profiling, cloud-based storage for expanded data retention, and integration with biometric authentication methods for enhanced security. Moreover, the incorporation of machine learning algorithms for continuous model refinement and adaptive access control policies further strengthens the system's capabilities. As technology evolves, the system can leverage advancements in facial recognition algorithms and hardware

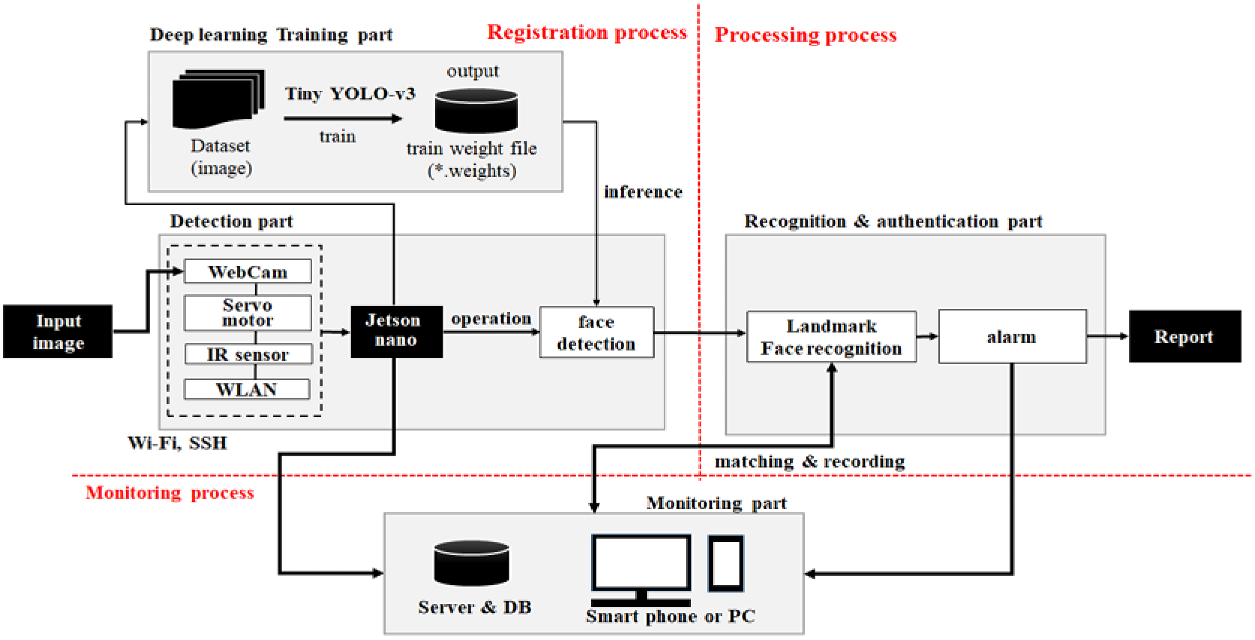


components to enhance accuracy, speed, and reliability. Additionally, the modular design facilitates easy integration with existing security systems, making it adaptable to diverse environments and use cases. Overall, the system's future enhancements aim to elevate security standards, improve user experience, and ensure compliance with evolving regulatory requirements.

III. PROPOSED SYSTEM

The system's overall structure, as proposed in this paper, is depicted in a block diagram (refer to Fig. 2). The system is primarily divided into two main components: registration and processing. These components encompass several sub-modules, namely (1) deep learning model training, (2) visitor face detection, (3) face recognition and access verification, and (4) monitoring. Within the deep learning model training phase, a learning model is developed using Tiny-YOLOv3, focusing on face features datasets categorized into seven classes. The visitor face detection segment constitutes the core of the face recognition security system. It integrates a webcam for CCTV functionality, a servo motor for environmental monitoring, and an infrared sensor for human body detection, all interconnected with the Jetson Nano board. Initially, the system captures real-time images from the webcam and employs the learning model to detect human faces. Upon confirming the presence of a human face, the system utilizes the landmark algorithm in the recognition and security module to identify detailed facial features. These identified features are then compared with the stored face data in the monitoring system's database. Access is granted if the detected face belongs to a registered member, triggering the cancellation of security measures. Conversely, if an unauthorized face is detected, an alarm notification is promptly issued to the user. Furthermore, the system logs comprehensive information, including facial data and headcount, which is wirelessly transmitted and stored in the database through the wireless LAN. This stored data allows for later analysis and verification of access events.

Figure 2. Structure diagram of the proposed system



DESCRIPTION OF THE COMPONENT:

1. **Deep Learning Model Training Component:** This component involves the derivation of a deep learning training model. Initially, images of family members and acquaintances are saved in a dataset. These images are then used to train the Tiny-YOLOv3 deep learning model.

2. **Video Data Collection Component:** This component is responsible for collecting visitor video data using the webcam CCTV.

3. **Control and Processing Components:** A database is established on the Jetson Nano board to store and process the collected video images.

4. **Monitoring and Post-processing:** The processed results are transmitted to the manager's smartphone for further processing.

The hardware configuration diagram depicts the system's operation for face detection, recognition, and log storage, utilizing NVIDIA's Jetson Nano board as the central system. This board is designed for artificial intelligence learning and features Ubuntu 18.04, a Quad-core ARM A57 CPU @1.43 GHz, a 128-core Maxwell GPU, and 4GB of 64-bit LPDDR4. The hardware components include Logitech's C270 HD Webcam for CCTV functionality, an HDMI LCD Touch Screen Monitor, a servo motor, a human body detection infrared sensor, and Logitech speakers. The hardware implementation involves connecting the Logitech webcam to the Jetson Nano board for CCTV purposes and integrating an infrared sensor for visitor detection. The system activates upon detecting a visitor through the infrared sensor. The CCTV captures images, which are then processed using the pre-trained deep learning model for face detection. Key facial features are identified and compared with stored face data to verify the visitor's identity. MariaDB is used to store video images and access information from CCTV. Internet connectivity enables visitor notifications and remote user monitoring. A wireless LAN card facilitates connectivity with the smart home. Additional authentication measures are implemented to enhance visitor face recognition accuracy. The system communicates with users via the touch screen, displaying messages and providing a keypad for inputting passwords when necessary. These components work cohesively to ensure efficient face detection, recognition, and access control within the system.

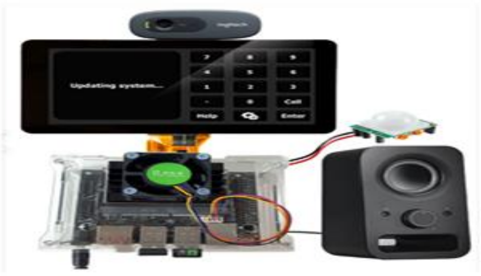


Figure 2. Hardware Design

IV. MODULES

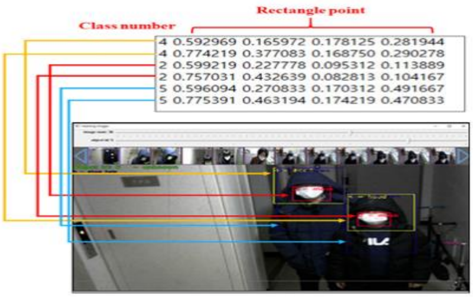
The system can be broken down into several modules, each serving a specific function within the overall framework of Visitor Authentication Model

A. Compilation of comprehensive Dataset:

we leveraged Google's capabilities to gather over 500 facial images exhibiting diverse characteristics from the web. In addition to this, recognizing the specific environmental context of the system's installation, we collected an additional 500 images depicting individuals entering and exiting the building. This meticulous dataset acquisition process ensured a wide range of facial attributes, including variations such as wearing eye glasses, masks, or caps, as well as diverse gender representations, age groups, and varying lateral angles of the face. The process involved utilizing CCTV footage, and Figure 3 illustrates the annotation of the visitor sample result screen. To facilitate seamless learning within the Jetson Nano platform, all images were standardized to a resolution of 400 x 320 pixels. Furthermore, considering the impact of environmental factors, the collected images underwent augmentation techniques such as rotation, inversion, and brightness adjustments to enhance robustness and adaptability. Out of the total 2,000 images gathered, 1,400 were meticulously curated for training purposes, 400 were allocated for validation, and the remaining 200 were reserved for rigorous testing, ensuring the dataset's reliability and efficacy in training the facial recognition model.

B. Classification Process

We utilize the YOLO Marker tool, an open-source labelling tool for YOLO (You Only Look Once), which is freely available on AlexeyAB's GitHub repository. This tool enables us to classify facial features with enhanced accuracy. Specifically, we select and classify seven crucial features within the facial region to ascertain that it indeed represents a human face. These features, outlined in Table 1, include key attributes such as index, class number, and bounding box coordinate values. Each image in the dataset is meticulously labelled, and the labelled data is stored in a .txt file, with a corresponding list file created for the entire image set. Each labelled item includes a class number assigned during labelling, followed by square coordinate values for each labelled class. For instance, consider the first value [4 0.5 0.1 0.2 0.3]: here, class name 4 signifies a specific feature, with the second and third values representing the x and y coordinates at the rectangle's centre point (0.5, 0.1). The fourth value denotes the width (0.2), and the last value indicates the height (0.3) of the labelled feature. It's important to note that the screen's width and height are standardized to 1.0 each for consistency in measurements. Additionally, class numbers 2 and 4 signify specific features such as masks and faces or heads, respectively, further enhancing the granularity and accuracy of the labelling process.



```
4 0.592969 0.165972 0.178125 0.281944
4 0.774219 0.377083 0.168750 0.290278
2 0.599219 0.227778 0.095312 0.113889
2 0.757031 0.432639 0.082813 0.104167
5 0.596094 0.270833 0.170312 0.491667
5 0.775391 0.463194 0.174219 0.470833
```

Class name	Class number
Eye_area	01
Mask	02
Eyebrow	03
Face	04

R_eye	05
L_eye	06

TABLE 1. Classification table for face features.

C. Training process

The training process for the Tiny-YOLOv3 model dedicated to face feature recognition was executed on the Jetson Nano platform. Illustrated in Figure 8 is the intricate architecture of the deep learning model tailored specifically for face detection, crafted through extensive learning from the dataset. The model's output is structured to pass through various stages, including input processing, Convolution at step 13, MaxPooling at step 6, Up-sampling at step 1, step 2 of detection, and final output. This architecture enables the model to effectively discern human faces by leveraging seven distinct features. Among these features, face detection holds paramount importance as it serves as the linchpin for recognizing humans amidst a plethora of extracted features, prioritizing attributes like eyes, nose, and mouth as the primary identifiers. Upon detecting four or more critical features, including the face feature, the model seamlessly transfers the extracted facial area data to the authentication system for further processing. Figure 4 showcases the inference results derived from the meticulously trained model, highlighting its proficiency in accurately identifying and delineating facial features within images.

Figure 3. visitor sample result screen Annotation

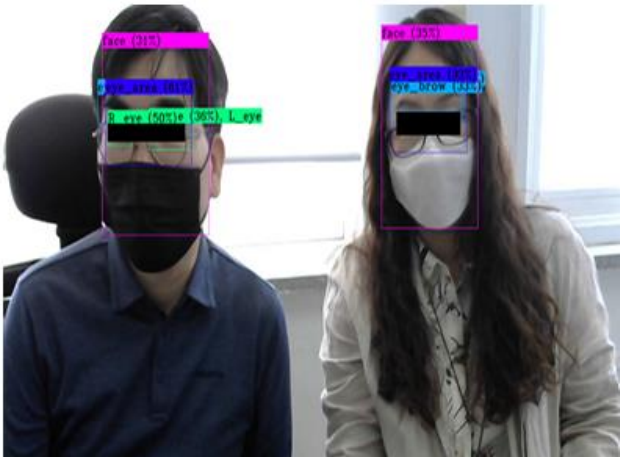


Figure 4. Inference Results

V. METHODOLOGY

A. USER AUTHENTICATION:

A MariaDB database is established on the Jetson Nano platform, dedicated to managing visitor access control efficiently. Within this database, a specialized table named pictureTBL is crafted to store crucial access-related information. This includes records of the visitor's image, their identity, the count of simultaneous visitors, and the timestamp of their visit. Additionally, to facilitate future post-processing and analysis, visit records are systematically registered in the visitorLogTBL, ensuring their separate and organized management. In cases where multiple visitors are detected, the system intelligently records the name and details of the foremost visitor for enhanced record-keeping and accessibility.

B. PROTOCOL:

Algorithm 1 is a pseudocode designed to identify visitors using a combination of the PIR sensor and the CCTV camera installed at the main entrance. If the system fails to recognize a visitor from a distance or when they are wearing a mask, it alerts the user to manually verify the visitor's face. During this process, the system captures a photograph of the visitor, creating a detailed visit history. Access to the premises is granted only upon successful recognition of a visitor's face, particularly if they are identified as a family member. In cases where initial identification is unsuccessful, additional verification steps are prompted.

[Visitor Recognition Steps]

1. The system distinguishes between family members, visiting guests, passersby at the front door, or couriers delivering goods. Recognition occurs only when a visitor remains in front of the door for a specific duration within the proximity sensor's range. Upon detection, the system activates from standby mode.

2. Utilizing the CCTV camera, the system captures a visitor's image for recognition, employing a deep learning-trained model to identify facial features.

3. Once the system detects a face, it captures an image centred on the face and stores it within the system for potential suspicious activity monitoring.

[Visitor Decision Steps]

1. Visitors may be couriers leaving deliveries or guests entering the premises. Guests or family members intending to enter ring the doorbell or input the door lock password. The system differentiates between business members and visitors based on wait times at the front door. Guests trigger the doorbell, while family members enter the door lock password.

2. The system verifies visitors by attempting to identify their faces. If facial recognition fails due to distance, it prompts the visitor to approach or delivers a message via speaker.

3. Upon successful face region detection, the visitor's face image is stored for record-keeping.

[Face Recognition Steps]

1. The system attempts face recognition using CCTV images.
2. If face recognition fails, the system checks for mask usage, requesting mask removal or delivering message via speaker.
3. If no mask is detected, the system reattempts face region detection and captures a new image.
4. Recognized faces are cross-referenced with the database to verify family member status.

[Post-Authentication Process]

1. Registered family members are authenticated simply, automatically opening the front door.
2. For guests, the visitor's face image is sent to the manager for confirmation. After verification, the manager may open the door. If face recognition fails, visitors can use the door lock password for authentication. The door lock system verifies passwords entered on the touchpad, opening the door when the correct password matches the stored one

Algorithm 1: Visitor Authentication Based on Face Recognition

```
Operate PIR Sensor: Detect visitors using the PIR sensor.
Operate CCTV: Capture visitor images using the CCTV camera.
1: while System is active do
2:   if Front door waiting time > Threshold then
3:     Guest is identified as a Visitor.
4:   else
5:     Guest is considered a passer-by.
6:     go to line 1
7:   end if
8:   while Visitor's face is not detected do
9:     Trigger Alarm("Please come closer.")
10:  end while
11:  Save photo of the Visitor.
12:  if Visitor's face is recognized then
13:    if Visitor is a Family member then
14:      Open the Door.
15:    else
16:      Trigger Alarm("Who are you?")
17:    end if
18:  else
19:    if Visitor is wearing a mask then
20:      Trigger Alarm("Please take down your mask.")
21:      go to line 11
22:    else
23:      go to line 8
24:    end if
25: end while
```

Algorithm 2: Visitor Counting and Anti-Shoulder Surfing

This algorithm calculates the number of visitors at the main door and includes measures to prevent shoulder surfing attacks. When face detection fails to authenticate a family member, additional authentication is required. However, this can lead to password exposure if there are multiple visitors. To avoid this, the system determines the number of visitors to prevent shoulder surfing. Here are the steps:

[Steps for CCTV Image Capturing]

1. The CCTV system detects faces in the image using a pre-trained deep learning model.
2. The system determines the number of faces in the captured image.

[For Two or More Visitors]

If there are two or more visitors, the system checks for shoulder surfing. It detects shoulder surfing by monitoring the position of individuals, their face size, and eye focus from behind. If a shoulder surfing attack is suspected, the system prompts the visitor to enter a lengthy or complex password.

[Steps for Password Input]

1. The visitor inputs the password on the touchpad.
2. If the entered password matches the one stored in the database, the front door opens.
3. Since the system relies on face recognition for authentication, it's susceptible to various attacks like using photographs of family members. To counter this, the system verifies the changing feature points on the face to distinguish between a real face and a photo. If it detects a photo, user authentication is blocked to prevent unauthorized access.

```
Algorithm 2: Prevention of Shoulder Surfing Attacks
Procedure: Password Input
1.  Capture photo from CCTV.
2.  Input detected face image of visitors.
3.  If the number of visitors is more than 2,
a.  Check for shoulder surfing attack using the shoulderSurfingAttack function.
b.  If a shoulder surfing attack is detected,
- Display "Please input a long and complex
            Password"
            - Input the password.
Procedure: shoulder Surfing Attack
1.  Determine the size of the other person's face (the person behind).
2.  Determine the position of their eye focus.
3.  Set peek to false initially.
4.  Introduce a delay of 300 milliseconds.
5.  If the size of the other person's face is greater than the specified size,
6.  Set peek to true.
7.  Else, if the difference between the other person's eye position and focus is less than the threshold value,
8.  Set peek to true.
9.  Return the value of peek.
```

C. EVALUATION ON THE ANALYSIS OF SECURITY FIELD:

The introduced system meets the outlined requirements in several ways. Firstly, it employs CCTV-based face recognition to determine the number of visitors. If this count decreases within a set timeframe, it identifies the individual as a passerby; otherwise, they are categorized as a visitor or potential intruder. Secondly, to address potential security risks, especially regarding password peeking, the system evaluates the behaviour of individuals behind the front visitor. If suspicious actions like attempting to view the password are detected through CCTV analysis, a warning message is displayed to the front visitor, prompting them to enter a more secure password. Thirdly, the system incorporates pre-registered family member data for comparison during visitor authentication. By analysing video images from the CCTV, it matches faces with registered ones to authenticate visitors. In cases where primary authentication isn't conclusive (e.g., due to low confidence scores), secondary authentication via a short password is required for access.

Fourthly, the system facilitates remote monitoring and access control via smartphones. Images captured by CCTV are transmitted to the manager's phone through the smart home service, enabling real-time intrusion detection. If primary authentication fails or the visitor is not registered as a family member, access is denied until verified by the manager remotely. The system's functionality includes real-time visitor authentication, remote access management, and robust security measures to deter unauthorized access. It ensures that only authenticated visitors gain entry, with comprehensive logging and monitoring features for enhanced security and convenience.

VII. ANALYSIS AND EXPERIMENTAL RESULTS OF THE PROPOSED SYSTEM

A. ANALYSIS OF USER AUTHENTICATION SYSTEM

When a visitor's face is recognized using CCTV imagery, a model trained on shape_predictor_81_face_landmarks.dat detects 81 feature points. The HoG feature acts as a face detector, and a linear classifier locates landmarks like the eyes, nose, mouth, chin line, and eyebrows. For user authentication, the faces of registered members or family members are pre-stored in the database. The vector values of their facial feature points are stored in the familyTbl table. This method then compares the feature point vector values of the visitor's face with those of the stored family members. If the threshold value falls within the authentication range, the authentication is successful, identifying the visitor as a family member if their face closely matches a registered member. However, due to potential variations in facial appearances, such as wearing masks, glasses, or other coverings, false acceptance (FAR) and false rejection (FRR) rates can occur. These rates indicate instances where a non-family member might be accepted or a family member might be rejected due to obscured or distorted facial features.

To mitigate these issues, Algorithm 3, implemented in Python using the Dlib library, extracts and saves 81 facial features from family photos. It captures and compares features from webcam images at the main door, verifying the visitor's identity against the stored family member features. In the proposed model, additional verification steps, like requesting a pass number or manual verification by an administrator, can be included. During face detection via webcam (CCTV), if two visitors are present, they are prompted to lower their masks for accurate feature extraction. This step ensures the system captures precise facial feature points when masks are removed, improving recognition accuracy.

C. EXPERIMENT RESULTS

The proposed system allows real-time processing of visitor face recognition using hardware such as Jetson Nano for efficient computation. It performs facial detection, eyebrow, mask, and eye detection, capturing images with various attributes like eyeglasses, masks, caps, and makeup. Additionally, it can determine gender, age, and the face's lateral angle to improve face detection accuracy and reduce false positives. In a similar setup, Raspberry Pi, a more affordable option compared to Jetson Nano, was also tested. The Raspberry Pi implementation utilized a model trained with Tiny-YOLOv3. It successfully detected three faces, with different numbers of facial features detected for each person. The authentication process, as depicted in the figures, involves identifying and comparing key features to verify if they match the registered person. The authentication process involves determining the existence of seven features, as shown in the figures, which indicate the probability of feature presence. The inference results and Frames Per Second (FPS) output of the algorithms, including YOLOv3 and Tiny-YOLOv3, are visualized to demonstrate their execution performance.

VI. CONCLUSION

Deep learning, a pivotal technology within artificial intelligence, is experiencing rapid growth, especially in voice and image recognition applications. Its impact extends to autonomous driving and crime prevention systems, shaping future industries. Within image recognition, advancements in Convolutional Neural Networks (CNNs) have led to various object detection algorithms, making deep learning a cornerstone in these domains. In this paper, we delve into different object detection algorithms, notably CNNs. Our system utilizes CCTV to detect and recognize a visitor's face, capturing 81 feature points to create a vector representation based on these features. Family members are pre-registered with their facial images. Our system is designed to open the front door only when a new visitor's recognized facial feature vector closely matches the stored database, falling within a specified threshold. We tested our proposed model on portable microprocessors capable of operating various sensors. While the Raspberry Pi and Jetson Nano were tested, the Raspberry Pi's performance was inadequate due to slow operation, whereas the Jetson Nano, with its built-in GPU, yielded the desired results. Specifically, in our model, YOLOv3 achieved an inference time of 2.4 FPS with 90.3% accuracy, while Tiny YOLOv3 achieved 6.5 FPS with 86.3% accuracy. Future endeavours include developing systems for face recognition in access control for shops or restaurants, automatic registration of frequent visitors, visitor count management, visit record storage, and access control protocols based on visitor profiles. Additionally, there's a need for an access authentication system that leverages facial recognition for easy access, offering a lower false detection rate compared to fingerprints or passwords

ACKNOWLEDEGMENT

We express our sincere gratitude to the Head of the Department, Dr. K. Geetha , for their invaluable support and guidance throughout the process of publishing our conference paper. We extend our heartfelt appreciation to our esteemed guide, Ms.M.Mathumitha, for their mentorship and expertise that greatly contributed to the success of our paper. Additionally, we would like to thank the Principal, Dr.X.Susan Christina, M.A.M college of Engineering and Technology(MAMCET), for their encouragement and endorsement of our research endeavors. This achievement would not have been possible without their unwavering support and encouragement

REFERENCES

[1] Ahmed, S. U., Khalid, H., Affan, M., Khan, T. A., & Ahmad, M. (2020). Smart Surveillance and Tracking System. In 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan.

[2] Aryah (2012). Biometric Visitors Management System. [ONLINE] Available at: http://www.aryah.net/VMS.htm. [Last Accessed 20 March 2014].

[3] Babanne, V., Mahajan, N. S., Sharma, R. L., & Gargate, P. P. (2019). Machine learning based Smart Surveillance System. In 2019 Third International conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India.

[4] Dai, J., et al., "Deformable convolutional networks," in Proc. of the IEEE Int. Conf. on Computer Vision, 2017.

[5] Deng, J., et al., "Arcface: Additive angular margin loss for deep face recognition," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019.

[6] Emgu CV Tutorial, 2013. Available at: http://www.emgu.com [Accessed 20 July 2013].

[7] Ferrari, C., Berretti, S., & Del Bimbo, A. (2018). Extended youtube faces: a dataset for heterogeneous open-set face identification.

[8] Hasam (2010). VMS 3000 Visitor Management System. [ONLINE] Available at: http://www.hasam.eu/en-us/products/identificationsystems/visitormanagementsystem.aspx. [Last Accessed 25 March 2014].

[9] HID (2013). EasyLobby® Secure Visitor Management (SVM™) Software. [ONLINE] Available at: http://www.hidglobal.com/products/software/easylobby/svm[Last Accessed 15 March 2014].

[10] King, D., "High quality face recognition with deep metric learning." URL: http://blog. dlib. net/2017/02/high-quality-face recognition-with-deep. html (2017).

[11] Lin, T.-Y., et al., "Feature pyramid networks for object detection," in Proc. of the IEEE Conf. on Computer Vision and

Pattern Recognition, 2017.

[12] Lin, Y., et al., "Mobiface: A novel dataset for mobile face tracking in the wild," in 2019 14th IEEE Int. Conf. on Automatic Face & Gesture Recognition (FG 2019).

[13] Liu, W., et al., "Sphereface: Deep hypersphere embedding for face recognition," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2017.

[14] Rabia Jafri, Hamid R. Arabnia, 2009, “A survey of face recognition techniques”, Journal of Information Processing Systems, Vol. 5, No 2. Available at http://www.cosy.sbg.ac.at/~uhl/face_recognition.pdf [Accessed 6 June 2014]

[15] Radhika. K. M, Shankar. M. Bakkannavar, Arjun. M. S, Samarth Bhaskar Bhat. Face Detection from CCTV Footage using OpenCV and Haar Cascade. Available at https://www.ijraset.com/best-journal/face-detection-from-cctv-footage-using-opencv-and-haar-cascade.

[16] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering.

[17] Tao, Y., Zongyang, Z., Jun, Z., Xinghua, C., & Fuqiang, Z. (2021). Low-altitude small-sized object detection using lightweight feature-enhanced convolutional neural network. Journal of Systems Engineering and Electronics, 32(4), 841-853.

[18] Wang, H., et al., "Cosface: Large margin cosine loss for deep face recognition," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018.

[19] Wojke, N., et al., "Simple online and realtime tracking with a deep association metric." In 2017 IEEE international conference on image processing (ICIP).

[20] Willelectronics, 2012. VISITOR MANAGEMENT SYSTEMS. [Online] Available at: http://www.willelectronics.com/visitor_management_systems.html [Accessed 20 April 2013].

[21] Zafeiriou, S., Zhang, C., & Zhang, Z. "A survey on face detection in the wild: past, present and future," Computer Vision and Image Understanding, vol. 138, pp. 1-24, 2015.