ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Face Detection Using SURF Algorithm

Usha Kamale

Department of ECE, MVSR Engineering college, Hyderabad, Telangana, India

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150400099

Received: 18 August 2025; Accepted: 27 August 2025; Published: 16 May 2026

ABSTRACT

Image Processing offers solutions to a broad range of real-world challenges. Security issues and theft have been on the rise for several decades. There has consistently been an absence of adequate security systems to ensure safety for both commercial and residential properties. Consequently, real-time surveillance has become essential. However, this necessitates high-resolution cameras and extensive storage systems to record and access the footage of the captured videos. In this study, an effort has been made utilizing a digital image processing approach that incorporates motion detection and face recognition techniques to minimize memory storage without compromising the integrity of the original image. This system aims to achieve surveillance without relying on high-end components and devices. The work is divided into three primary components: motion detection, face detection and ultimately face recognition. The reliability and efficiency of the system can be enhanced by improving its accuracy and speed. This system can be utilized by consumer markets for the surveillance of their properties. The industrial sector can adopt this method to bolster security and to ascertain whether the detected individual is an employee. This approach can be applied in apartments, home automation systems, R&D test units, restaurants and various other commercial environments.

Keywords - Video Processing, Feature extraction, SURF algorithm, Face recognition, Surveillance, MTCNN

INTRODUCTION

The act of keeping an eye on or safeguarding a person, place of business, property etc. is known as surveillance. Since there are a lot of crimes in our society and inadequate surveillance techniques have made it harder to identity the true offender, security worries have grown over time. Many small business owners and homeowners have made significant investments in new and enhanced surveillance systems as a result of the rise in crime rates, rendering the older systems obsolete and useless. Nowadays, the same issue—security—occurs in many places, particularly cities, worldwide, and buyers are left perplexed by the never-ending search for suitable items. A lot of money is spent on hiring a large number of people to defend a location or assure public safety. The globe is also growing increasingly concerned about privacy and attempting to apply various strategies to incorporate privacy into their daily lives by avoiding intrusions.

Surveillance systems are strategically positioned by positioning a network of video cameras in the designated area and recording the events as they happen in order to detect an intruder entering a secured area or to monitor anything. Tis recorded video can be saved for later use or retrieved and seen on a monitor. Closed circuit television cameras are one type of security system that is frequently employed. These can be placed in residences close to the parking lot or utilized in public areas near poles or traffic signals. CCTV is widely used all around the world. However, the hard disk needs to have a lot of storage space in order to hold the recorded videos. Because of this, there is a lot of interest in the study being done on how to improve security with the newest technologies.

The way to do this is to upgrade just the most important components of the current system while simultaneously improving the image processing tools. This study examines the creation of a commercially viable smart surveillance system that uses face recognition and motion detection to identify the intruder's details. This system's benefit is that it minimizes memory storage by only storing video when motion is detected.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

LITERATURE SURVEY

This is our initial concept for creating a smart surveillance system that allows for citizen participation and data analysis for improved decision making, following security issues that have taken grip of our life. Together with the ability to identify and recognize faces, the smart surveillance system is straightforward to install and upgrade at the software level.

By building a database at home, a more privacyfocused version of the monitoring system that is cloud-centric can be implemented globally. The future potential, important technologies and applications that are anticipated to propel image and video processing research are all capitalized upon in this work. However, a solid basis for our work is given, combining the fundamentals and uses of face recognition, motion detection, face detection with Multi-Task Cascaded Convolution Neural Networks (MTCNN),and background subtraction into a single entity. Because it instills in citizens a sense of accountability and security, it is highly intriguing.

Analysis of Previous Work

As discussed in research papers Ref. [1] to Ref. [10] by researchers in the past, there are many merits and demerits in each proposed system by respective authors. The paper on "A deep learning approach to building an intelligent video surveillance system" by jiexu Ref. [1] is the base paper for the work carried out. Ref. [5] uses jetson tx2 board but does not contain the flexibility and upgradability integrated into the system, it also does not support face recognition.

Merits and Demerits observed in Ref. [1-11]

- Intimation of intruder using website but is lagging and easy to hack due to lack of proper encryption.
- Ineffective in some cases and poor output when used with cost effective systems.
- Previously designed systems cannot detect faces when viewed from other angles except facing towards the camera.

METHODOLOGY DESCRIPTION

This section provides a block schematic of the entire proposed work. In the sections that follow, each block is thoroughly discussed. The proposed system lessens the load on Government and the trouble of identifying the individual. Even when viewed from perspectives other than directly at the camera, this technology is able to identify faces.

Block Diagram of the proposed system

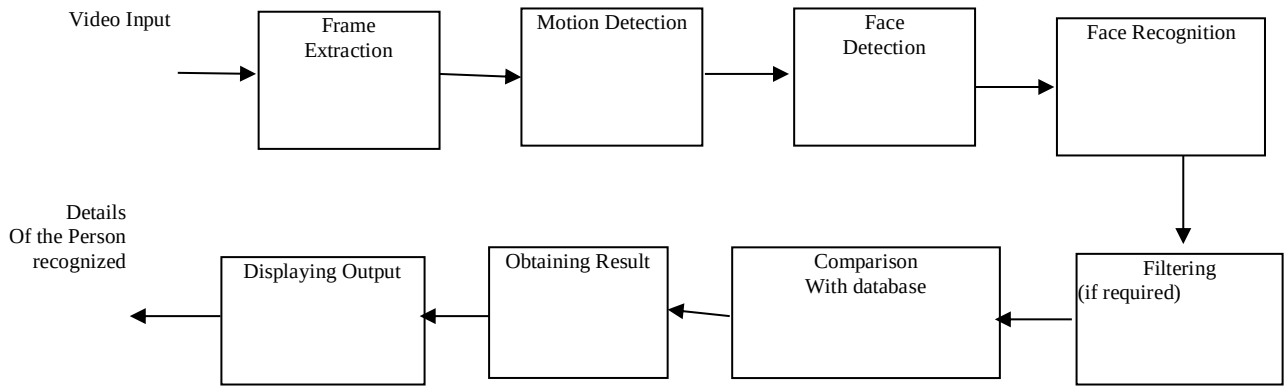

Fig.1: Block Diagram

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Frame Extraction

Firstly, a video of duration 40 seconds is taken as input. The video is read and divided into number of still images also known as frames. These extracted frames are then stored for future use. While running the program, 600 frames are extracted from the input video.

Fig.2: A glimpse of extracted frames

Motion Detection

A motion detection analysis is performed on the retrieved frames. The background subtraction approach is taken into consideration for this procedure.

A method of image processing called background subtraction is used to identify an image's foreground by subtracting the background. Masking, a subset of picture segmentation was segmentation, is used in this method. Masking, a subset of picture segmentation is used in this method. Threshold segmentation was employed here. This work uses the Frame Difference method in the background subtraction technique. This method considers two frames, calculates the absolute difference between them and then uses a threshold value to create a binary ask that is clear and accurate.

Face Detection

The output of the background subtraction method is taken and cropped to obtain the face region that needs further focus in order to detect faces. If necessary, this cropped face image may be filtered. Our study makes use of a deep learning technique called Multi Task Cascaded Convolution Neural Networks to detect faces. This three stage Neural Network technique provides precise face detection results.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Filtering

This block is employed when the face being recognized is blurry or unclear. Depending on the image identified from the output of the previous stage, it entails filtering operations like de-blurring, boosting, sharpening or contrasting.

Comparison and Output Display

In this step, the images in the database folder are compared with the output from previous step. The feature matching method is used for the comparison. The features of the matched individual are shown to the user as the output if the photographs after comparison are comparable to the degree that exceeds the specified threshold value. If the images do not match with the database, then it is displayed as 'NO MATCH' to the user as output.

Algorithms

In this section, the algorithm required for processing the frames of the video footages obtained. The various algorithms discussed in detail in this section are Motion Detection, Face Detection and Face Recognition. The recognized image obtained from the recognition process is also compared with the available database to get whether the output is matched or not.

Algorithm for Motion Detection Process

For Motion Detection, **Background Subtraction** technique is implemented. It is an approach where in the foreground image is separated from the background in a series of video frames. In Background Subtraction technique, Frame Difference Method is opted for this work. In this method the two frames from the extracted frames are considered and the absolute difference between the both frames is taken for the process. This process is carried out using difference using two images and then connecting the adjacent pixels using the adjacency and connectivity concept. This is a simple but an effective approach used by many systems. It is widely used in videos taken by static camera to generate a clean background image of the filmed scene or moving foreground objects.

The background subtraction process is loaded with the ideal condition that is visibly present most of the time. The image is then compared with multiple frames for reference. As long as there is difference between the frames, it means that there is a moment or motion detected and the frames will be saved. If there is no difference between the frames, it implies that there is no motion detected. Later the frame gets deleted from the storage.

Background Subtraction process involves the following steps-

1. Pre-processing
2. Background modeling
3. Foreground Detection
4. Data validation
5. Model Update

Pre-processing

In most computer vision systems, smoothing is used in processing to reduce high frequency noise from a digital image. It is also used to remove transient environmental noise like rain and snow captured in outdoor camera.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

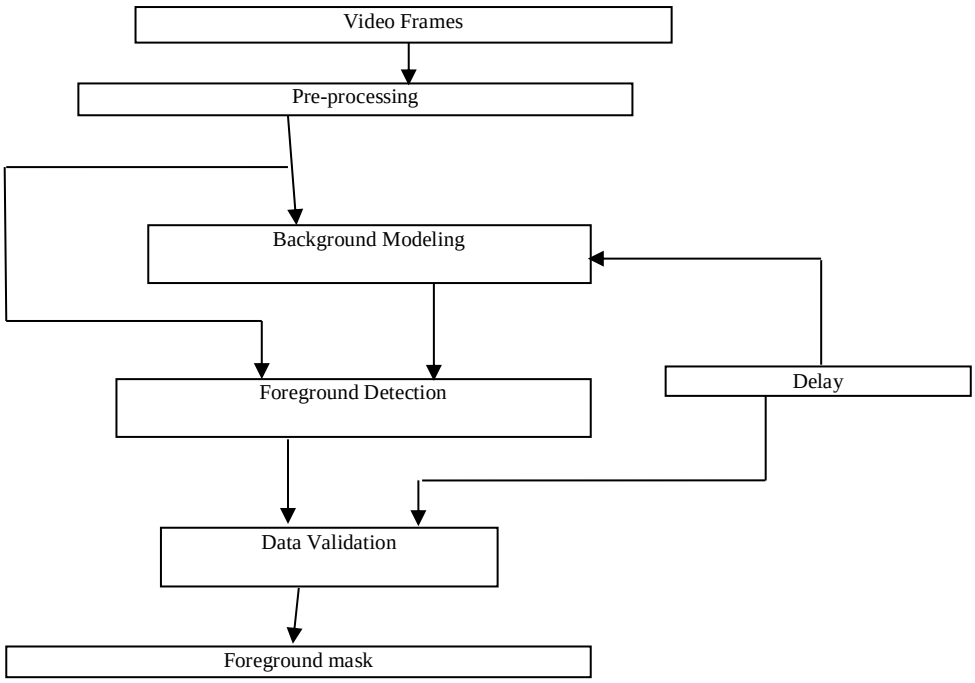

Fig.3: Block diagram for Background Subtraction Process

Background Modeling

Background modeling is at the heart of any background subtraction algorithm. A background model should be robust against environmental changes in the background, but sensitive enough to identify all moving objects of interest. In background modeling techniques are classified into two categories: Non-recursive and recursive.

A non-recursive technique uses a sliding-window approach for background estimation. Non-recursive techniques are highly adaptive as they do not depend on the history beyond those frames stored in the buffer. Frame difference, median filter, mean filter are some examples of Non recursive algorithms.

For recursive techniques, it does not maintain a buffer for background estimation. Instead, they recursively update a single background model based on each input frame. As a result, input frames from distant past could have an effect on the current background model. Compared with non-recursive techniques, recursive techniques require less storage, but any error in the background model can stay for a much longer period of time. Some examples of algorithms found in this category are Approximated median filter, Kalman filter and Mixture of Gaussians.

Here, in this paper, Frame Difference method which is a non-recursive technique of Background subtraction.

Foreground Detection

Foreground detection compares the input video frame with the background model and identifies candidate foreground pixels from the input frame.

Data Validation

This phase is sometimes referred to as the post-processing phase of the foreground mask (pixels).In this phase the candidate mask is examined, it is a detection algorithm where decisions are made independently at each pixel with isolated foreground pixels, it detects the holes in the middle of connected foreground components and jagged boundaries.

Therefore, in short the process for background subtraction is pre-processing of two images that are loaded in the algorithm. During the preprocessing step, the images are reshaped and enhanced for proper detection of the changes in the frames. Then the two images are subtracted from which we can detect the changes in the frames

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

intensities. To avoid unwanted detection of frame moment, a threshold is set to segment the image. This process is continued using the morphological filtering. During the morphological processing a collection of non-linear operations related to the shape, size and features are extracted. These features only detect the adjacent connecting pixels. These are the detected moments in the image. This approach can be understood by the following diagram.

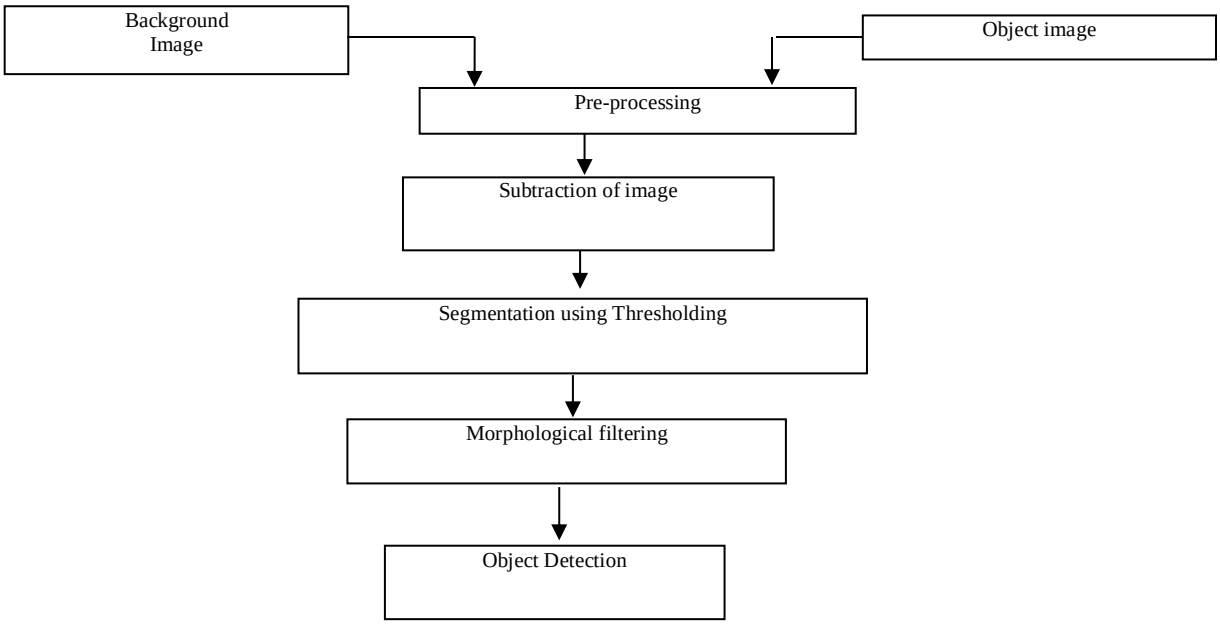

Fig. 4: Block diagram of Background Subtraction Algorithm

Algorithm for Face Detection Process

The process of detecting multiple of a single face present in an image is called as face detection. The process of detecting a face in out model is **MTCNN** which stands for **Multi Task Cascaded Neural Networks**. It is useful as it can run in real time on small devices. This system enables the users to implement it completely without using too much of data to run the algorithm. This is a neural network model which is used to detect faces and facial landmarks in an image. The accuracy of MTCNN is the strongest when compared to many of the models being used today for practical applications. MTCNN is a deep learning neural network which consists of 3 neural networks connected in a cascaded form. The three layers in a MTCNN system is as shown below.

Stage 1:

P-net: It stands for Proposal network. In this stage, it produces a candidate windows used for detection of features of the person by using a shallow convolutional process. It creates multiple frames which scans through the entire image starting from the top left corner and eventually progressing towards the bottom right corner. It is a fully connected CNN (Convolutional Neural Network).

Stage 2:

R-net: It stands for Refinement network. This stage is used to reject as many non-faces windows as possible. The neural network used here is complex and deeper compared to the previous stage. It is named as refinement stage as it refines the faces that have to be detected.

Stage 3:

O-net: It stands for Output network. This stage uses a complicated network for detection of faces and refinement in the image. It is the final stage of the process and as the name suggests it outputs the facial landmark position detecting a face from the given image or a video input.

A Convolutional Neural Network (CNN) is a deep learning network algorithm which can take in inputs, assign

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

weights and biases to the image. The aspects in an image are differentiated by using the weights and biases in an image. These images require very little pre-processing steps.

These neural networks are based on the connectivity pattern of the neurons in a human body. These neurons in the CNN algorithm are used to stimulate responses when set constraints are met. These constraints are known as receptive field. This process is similar to that of the human neurons behavior. Multiple fields such as the one mentioned above are connected together and overlapped to cover the entire visual area in the image.

This method of CNN is used because it can capture the spatial and temporal dependencies in an image throughout the application of relevant filters. The CNN is an algorithm which performs an excellent task of reducing the images into a form which is easier to process, without losing features which are important for accurate prediction.

Algorithm for Face Recognition Process

The process of face recognition is carried out using **Speeded Up Robust Features (SURF)** extraction method is a local feature extraction method. This is one of the most popular methods used for feature extraction, feature detection, object detection and 3D reconstruction. It works on the principle of detection and description of local features in digital images. These descriptions are used for defining quantitative information for the detection of features in an image. The algorithm for SURF consists of 3 steps–

1. Interest point detection
2. Local neighborhood description
3. Matching

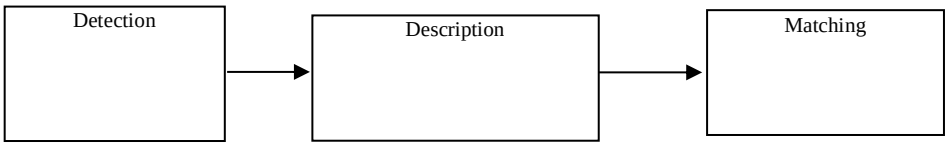

Fig. 5: Steps in SURF extraction method

Detection

SURF uses square shaped filters such as Gaussian filters for the detection of points of interest. The reason for selection of a square filter is because it is faster and easier to compute compared to other processes.

Descriptor

Descriptor is used to provide a unique and improved detection of an image feature. The descriptors are used for the point of interest at each and every point of interest identified in an image. A reproducible orientation of the image is first created based on the orientation of information along a circular or closed region along the points of interest. Then the SURF descriptor is used to extract the features form the image.

Matching

The matching process is then used to match the features from the inserted first image with the feature or object that we want to detect.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

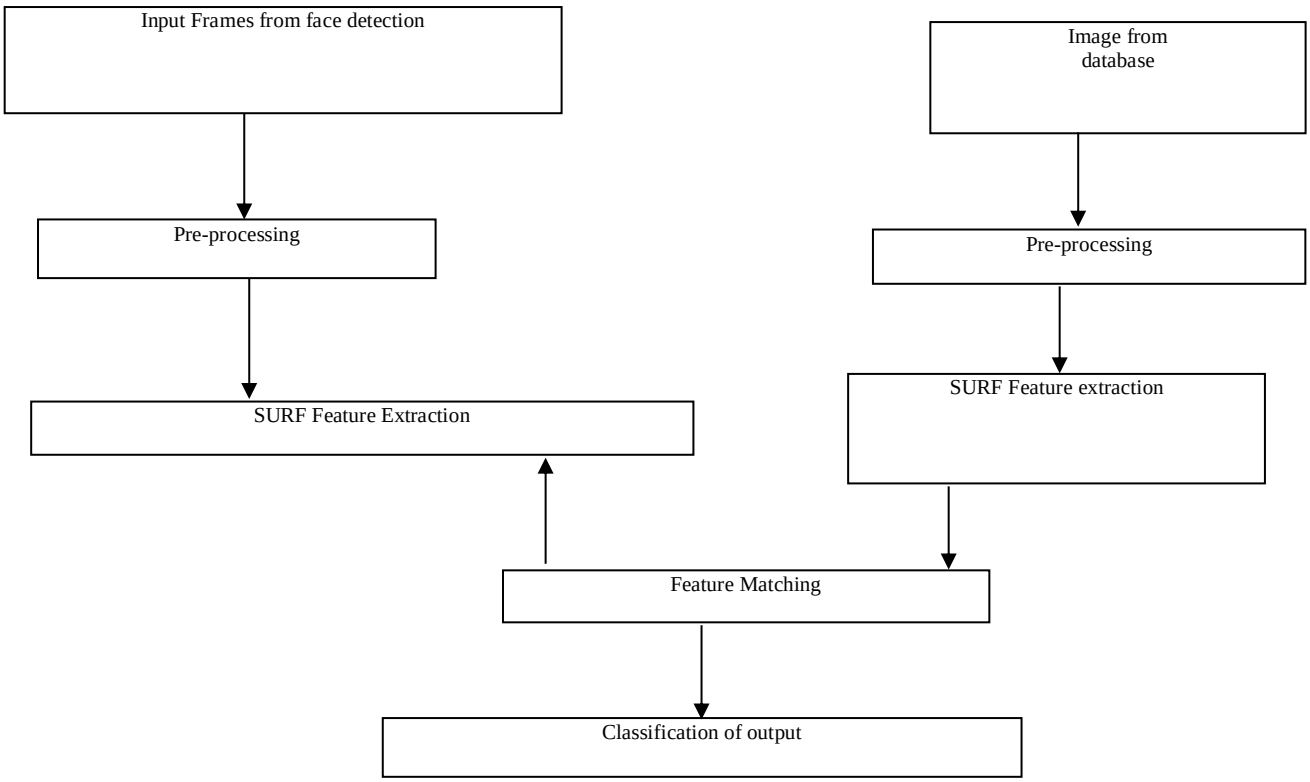

Fig. 6: Block Diagram of SURF Feature extraction

The above processes will happen in a successive fashion. This image input for this process is given from the multi task cascade neural network. This input image then undergoes the above 3 steps until it reaches the step of comparison with a database.

After the detection of the motion in the frames, the process is then passed to the feature extraction method this method uses the surf process which is an effective method for detection of features in the image. The input frames from the face detection process are passed on to the feature extraction process during this process, the images goes through pre-processing step, this step resizes, filters and enhances the input images from both the database images and the input image from the face detection algorithm.

This is then passed through the SURF algorithm, in this algorithm, the regions are detected and a circular region is analyzed, after the analysis the points are marked which are detected from the two input images. The input images are then compared and the points are selected according to the bounding region.

Comparison with Database

The input image from the SURF extraction is matched with the database images present. These images are looped until the correct image is found for matching an image. If the matching is passed, the surf process will stop and if the image does not match, the process will keep continuing until an image is matched or the complete database images have been completed for comparing. If there is no image that is matched with the images in the database, the output shown will be a message saying “No Match Found” and if there is a match confirmed with an image in the database, the output shown will be the details of the image matched with a message saying “Match Found”.

RESULTS

From the given video input 600 frames are extracted. The Frame Extraction, Motion Detection, Face Detection, Face Recognition and comparison has been carried out using MATLAB software. The results obtained from this are discussed below in detail.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

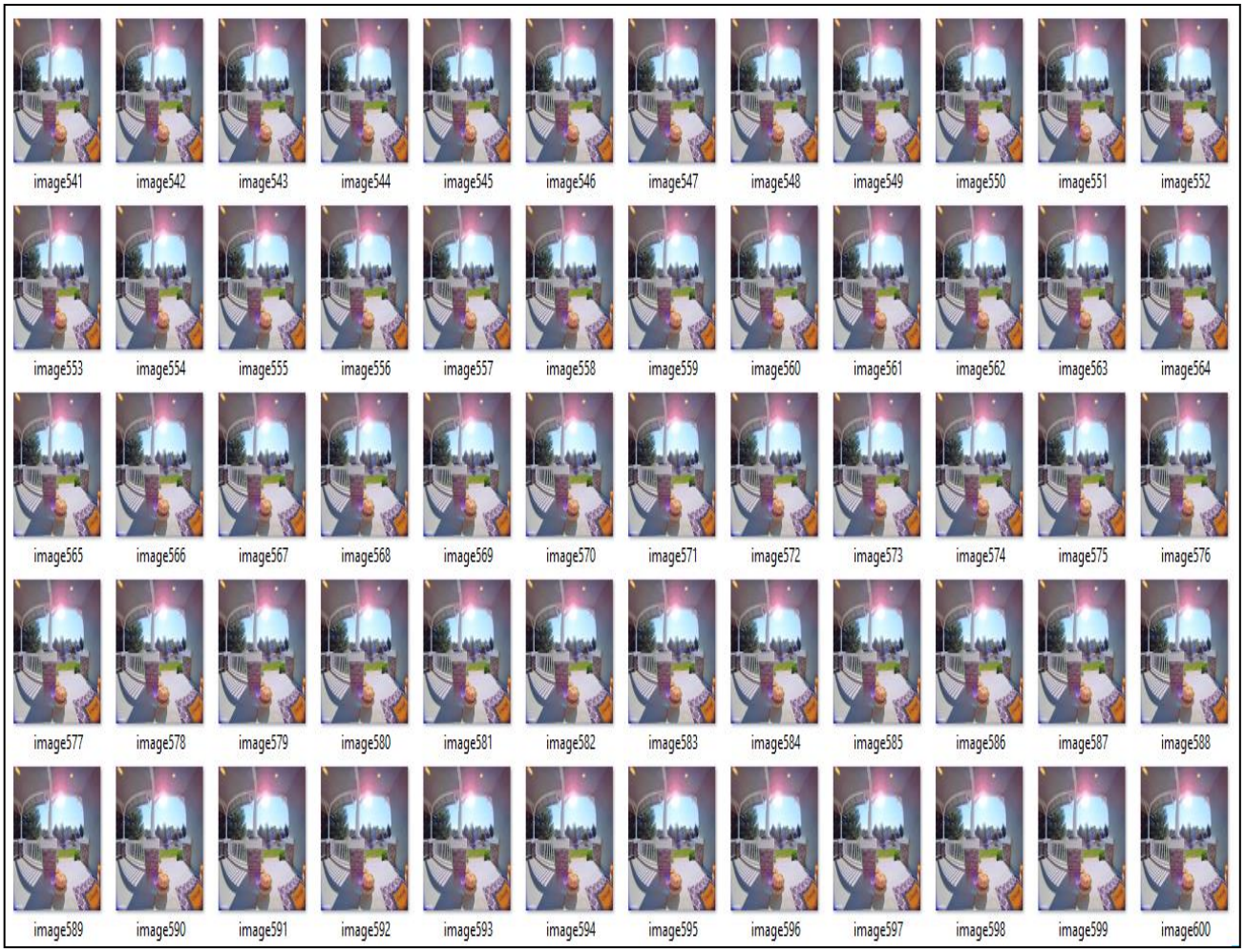

Fig. 7: Frame 541 to Frame 600

Fig. 8: Background Image Frame

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Fig. 9: Scene Image Frame

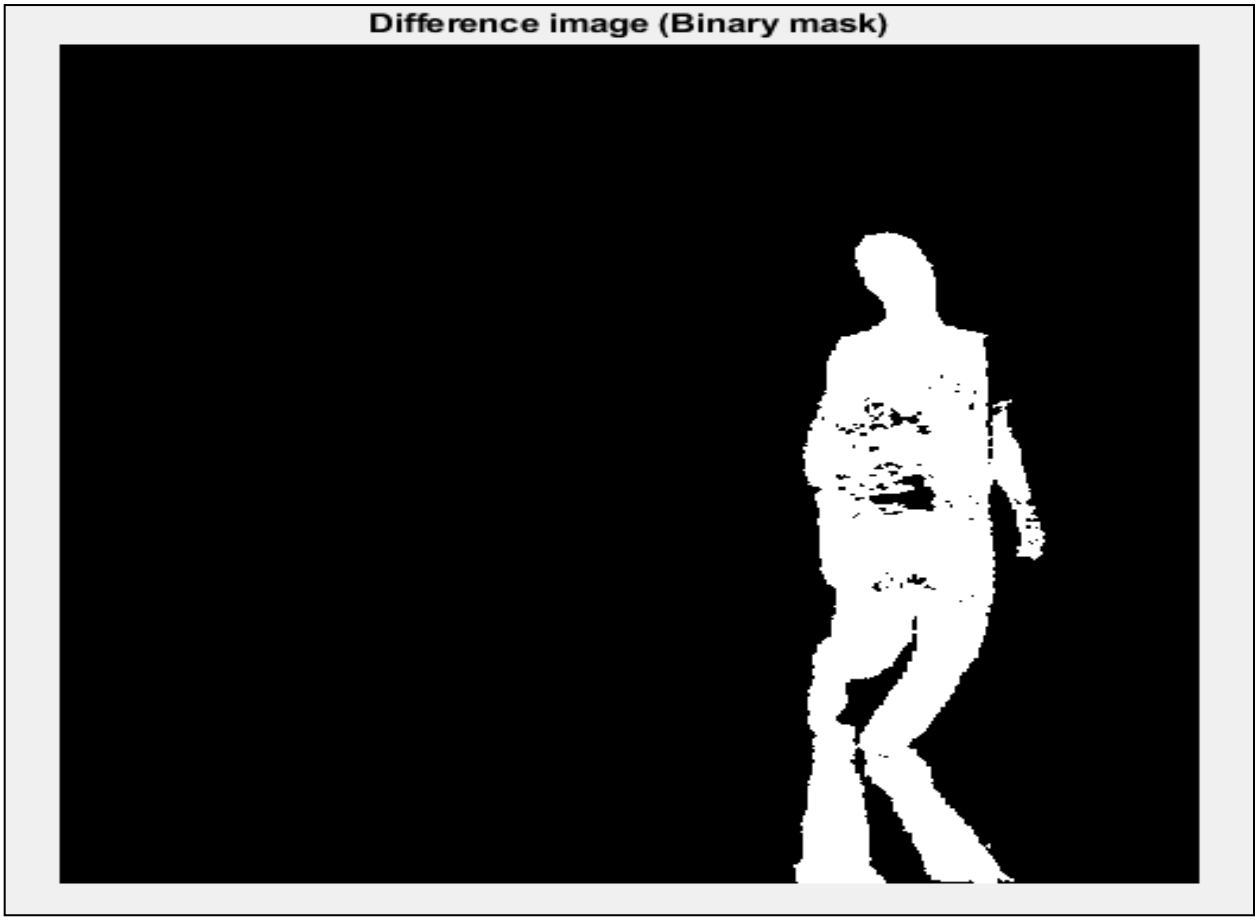

Fig. 10: Binary Mask of the Scene Image Frame

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

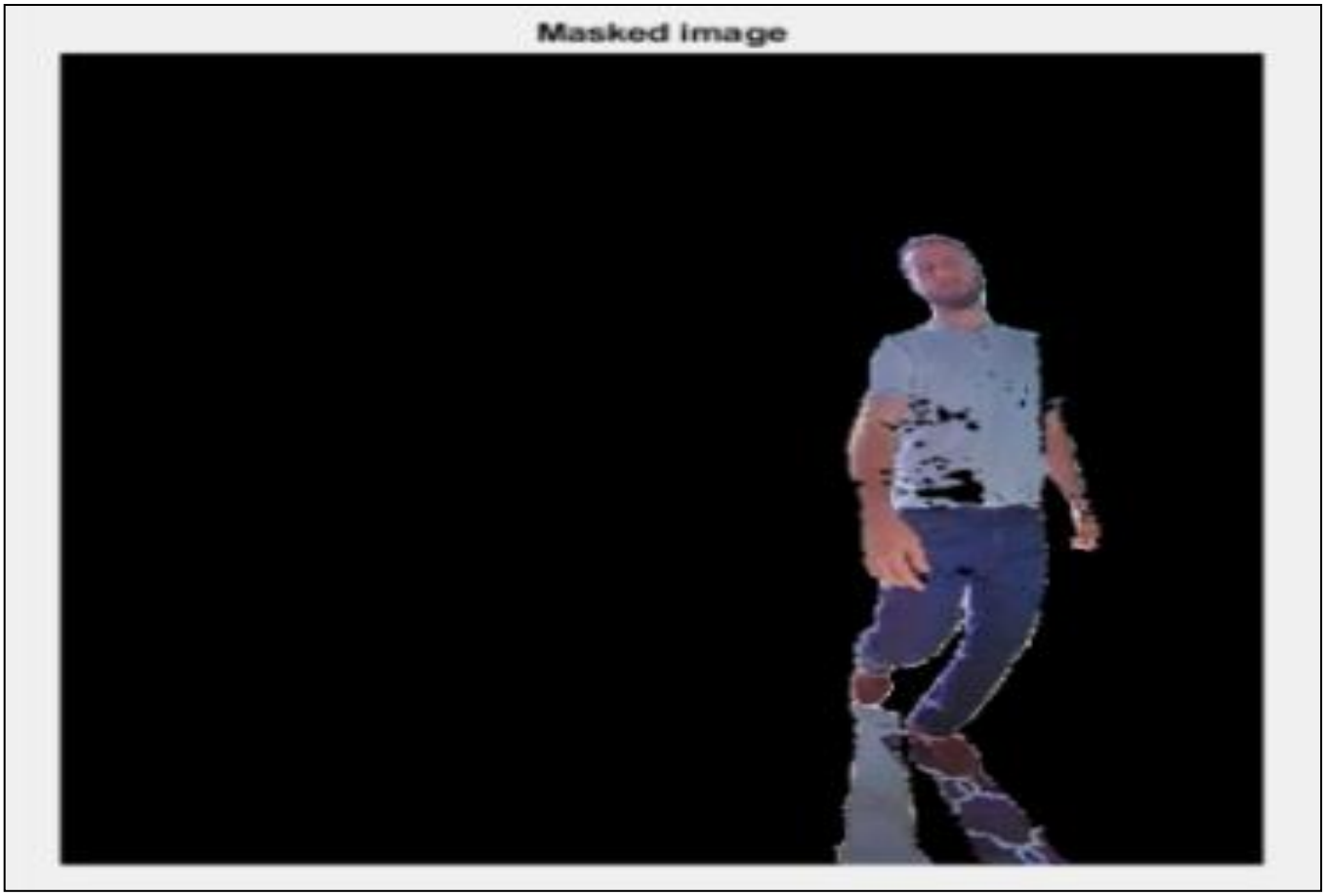

Fig. 11: Masked Image using Binary Mask

Fig.12: Image representing the detected face

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Fig. 13: Cropped Image of the face detected

Fig. 14: 100 strongest SURF Features from the cropped face image (Box image)

There are two cases in the output. The first case occurs when the recognized face is not matched with the database giving the output as 'No Match'. The second case happens when the face recognized image matches with the image in the database. The output of this case will be 'Match Found'.

Case 1- No Match

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Here the output is 'No Match' because the threshold value set for matching is greater than 2 features matching with the cropped image.

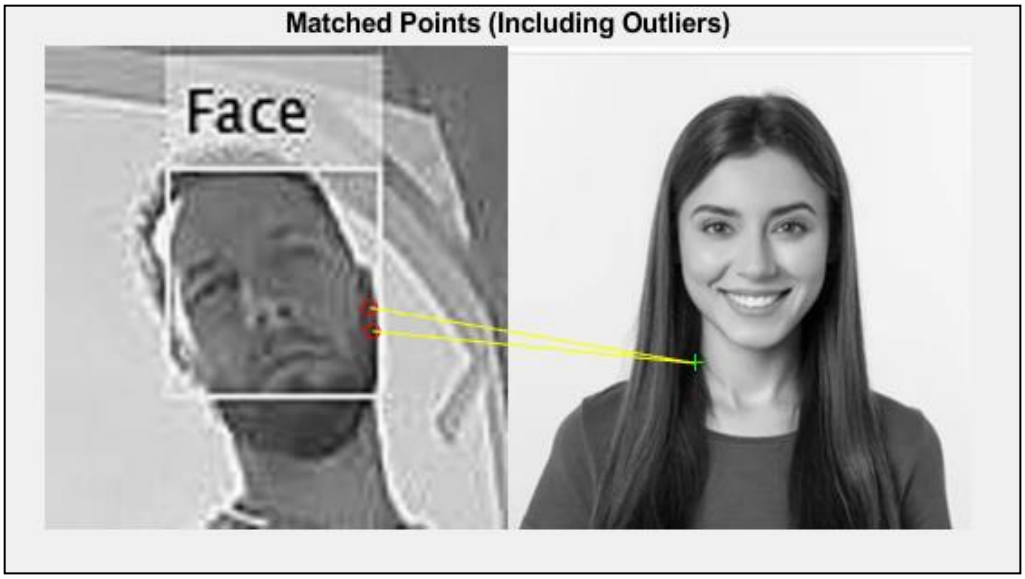

Fig. 15: Matched points between Scene Image and Box Image

Here the output is No Match because there are no matching features with the cropped image.

Fig.16: Matched Points between Database Image and Cropped Image

Case 2- Match Found

In this case, the match is found as the matching features are clearly visible as indicated by the lines drawn in the figure shown below.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

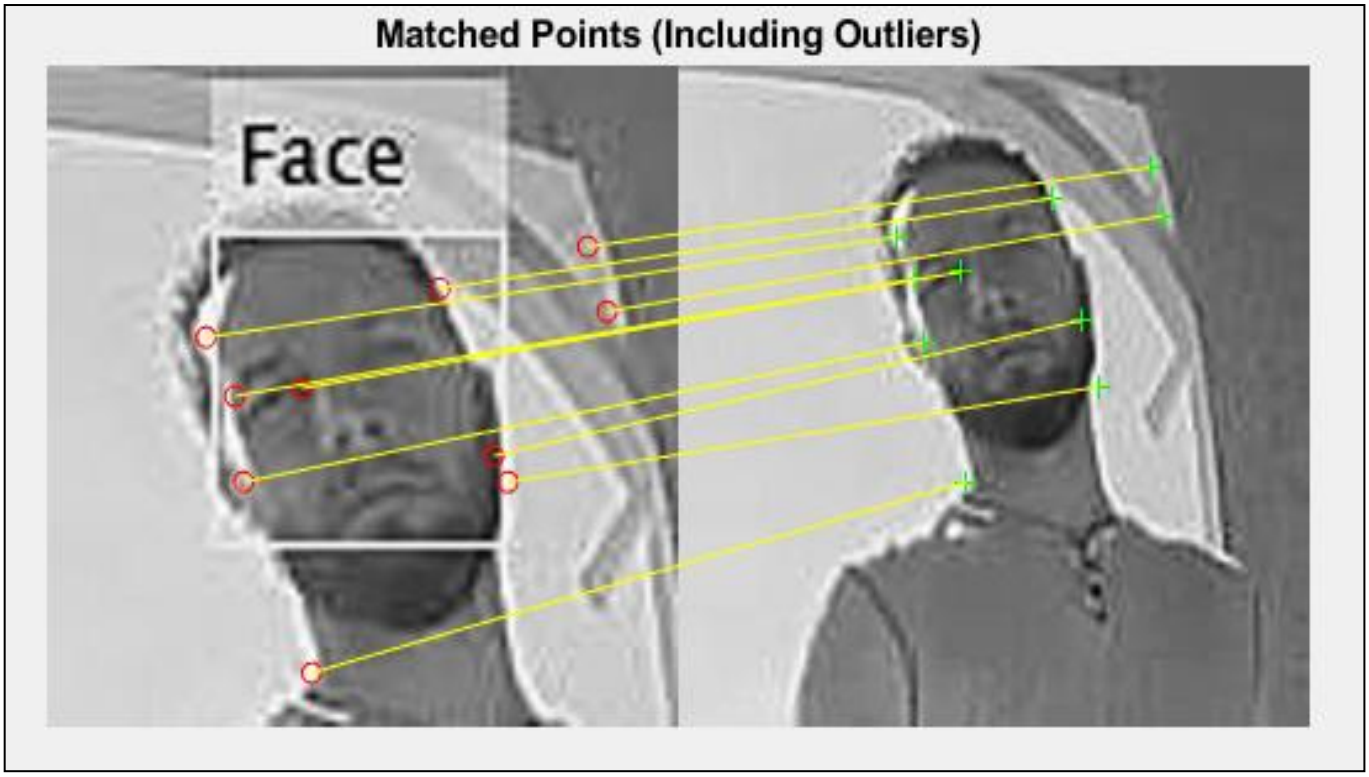

Fig.17: Matched points between database Image and the cropped Image (Including the Outliers)

Fig. 18 shows the facial regions (inliers) of the image.

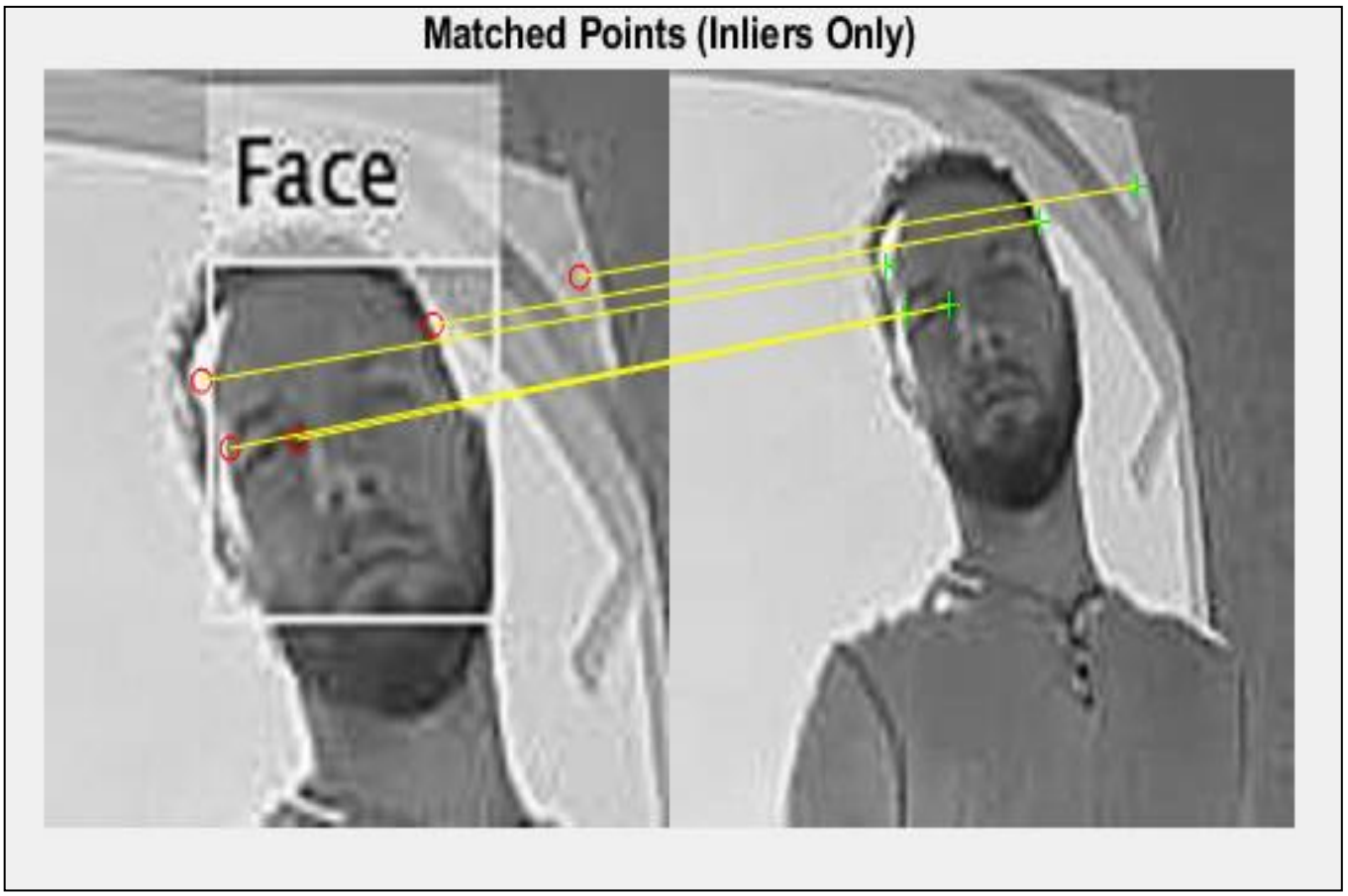

Fig.18: Matched points between database Image and the cropped Image (with only Inliers)

The next image shows the detected image indicated by the box drawn over the matched features.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Fig. 19: Detected image

CONCLUSIONS

This method is very effective in surveil the given regions. Rather than using conventional methods, here priority system is used to keep the regions under surveillance protected without upgrading the old systems. Secondly, using highly advanced software, this work has been implemented and the outputs have been observed. Thus in conclusion, the work has been successfully designed and tested and obtained the required results in an accurate and precise manner. This work is mainly intended to design and implement existing surveillance equipment without changing the components. This work can be extended by installing web-application and connecting to a mass database system for more analysis on data collection. The database can be increased so as to get the required output with more accuracy to recognize or match the features with different images.

REFERENCES

1. JieXu, “A deep learning approach to building an intelligent video surveillance system” , Multimedia tools and applications, Vol 80, pp 5495-5515, 2021.
2. AhireUpasan, BagulManisha, GawaliMohini, KhairnarPradnya, “Real Time Security System using Human Motion Detection”, IJCSMC, Vol. 4, Issue. 11, November 2015, pg.245 – 250.
3. Muhammad Awais, Muhammad JavedIqbal, Iftikhar Ahmad, Madini O. Alassafi, Rayed Alghamdi, Mohammad Basheri, and Muhammad Waqas, “Real-Time Surveillance Through Face Recognition Using HOG and Feedforward Neural Networks”, IEEE Access Volume 7, 2019.
4. Vivek srivastava, Ekta Chaturvedi
5. RajendraKachhawa, Raj Kumar Jain, ‘‘Security System and Surveillance using Real Time Object Tracking and Multiple Cameras’’ , Advanced Materials Research Vols. 403-408 (2012) pp 4968-4973 .
6. E. Jose, G. M., M. T. P. Haridas and M. H. Supriya, "Face Recognition based Surveillance System Using FaceNet and MTCNN on Jetson TX2," *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 2019, pp.608-613, doi: 10.1109/ICACCS.2019.8728466.
7. S. S. Thomas, S. Gupta and V. K. Subramanian, "Smart surveillance based on video summarization", *2017 IEEE Region 10 Symposium (TENSYMP)*, pp. 1-5, 2017.
8. Savath and Supavadee, "Real-Time Multiple Face Recognition using Deep Learning on Embedded GPU System", *Proceedings APSIPA Annual Summit and Conference 2018*, pp. 1318-1324, Nov. 2018.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

9. M. Ma and J. Wang, "Multi-View Face Detection and Landmark Localization Based on MTCNN", *2018 Chinese Automation Congress (CAC)*, pp. 4200-4205, 2018.
10. D. Meena and R. Sharan, "An approach to face detection and recognition", *Proc. Int. Conf. Recent Adv. Innov. Eng. (ICRAIE)*, pp. 1-6, Dec. 2016.
11. B. S. Satari, N. A. A. Rahman and Z. M. Z. Abidin, "Face recognition for security efficiency in managing and monitoring visitors of an organization", *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, pp. 95-101, Aug. 2014.
12. K. Vikram and S. Padmavathi, "Facial parts detection using Viola Jones algorithm", *Proc. 4th Int. Conf. Adv. Comput.Commun. Syst. (ICACCS)*, pp. 1-4, Jan. 2017.