Implementation of surveillance system through face recognition using HOG algorithm

Jung Kyu Park[1], Ji Won Yoon[*2] and Jung-Won Kim[*3]

[1]Professor, Department of Computer Engineering, Daejin University, Pocheon-si, 11159, Korea
[*2]Professor, CHARIS College of Liberal Arts, Changshin University, Changwon-si, 51352, Korea
[*3]Professor, Department of Beauty Arts, Youngsan University, Busan-si, 50510, Korea
Jkpark.@daejin.ac.kr[1], jwyoon.@cs.ac.kr[*2], jw__hada@ysu.ac.kr[*3]
*: jwyoon.@cs.ac.kr and jw__hada@ysu.ac.kr

How to cite this article: Jung Kyu Park, Ji Won Yoon, Jung-Won Kim (2024). Implementation of surveillance system through face recognition using HOG algorithm *Library Progress International,* 44(2), 639-674.

ABSTRACT
Recent years have seen a sharp rise in crime, raising concerns about neighborhood safety among the general public. A number of security technologies, such as digital door locks, closed-circuit televisions (cctvs), and alarm systems, have been developed and implemented. Even though most buildings have cctv installed, manually monitoring the footage takes a lot of manpower. The hog (histogram of oriented gradients) algorithm is widely used for object detection in computer vision. In particular, it shows good performance in face recognition and motion detection. In addition to identifying the user, the proposed system allows for mobile engagement, allowing the user to see the video, set off the alarm, and even receive notifications when an unknown person approaches the residence. The system proposed in the paper seeks to improve the performance of the current surveillance system by applying additional functions. The system was implemented to recognize people through real-time monitoring and send warnings if they are not registered users. Additionally, the system can be used to quickly report to the police.

Keywords: Computer Vision, Face Recognition, Hog Algorithm, Motion Detection, Surveillance Platform

1. INTRODUCTION

Numerous anti-theft systems have been created to reduce the likelihood of burglary incidents. Anti-burglary products include things like wireless security cameras, home alarm systems, motion-activated lighting, deadbolt locks with keypads, window or door alarms, and more. Surveillance cameras are widely used not only in companies but also at home, and are widely used to prevent theft. A typical home surveillance system includes surveillance cameras, motion sensors, sound sensors, and alarm notification functions. Most modern home alarm systems come with a wireless security camera as part of the package. The mobile application that lets users communicate with the system via their phone is another intriguing element of the most recent home alarm system. which enable the user to utilize a mobile application to set off an alert, watch the security camera, check the status of a sensor, and more [1-4].

Existing security cameras and home alarm systems do offer enough safety and features. But owning one of these systems comes with a hefty price tag. According to a recent survey, surveillance systems include object motion detection, sound detection, remote real-time monitoring, mobile phones, and recording functions. Additionally, the basic system starts at approximately $500 and can become more expensive depending on the number of cameras or storage capacity [5, 6].

Aside from that, modern surveillance cameras simply have the ability to record footage for later use and to monitor the area. To make sure the dwelling space is safe, though, the user will need to keep an eye on it around-the-clock. If they don't, they might miss the time when a burglar tries to get in and be unable to take quick action. In order to address this problem, security businesses provide their customers with a monitoring service that entails having professionals watch over their neighborhood around-the-clock and notify them right away if they suspect a break-in. The consumer will have to pay the security business a monthly charge in order to subscribe to the service. The user's privacy is this solution's primary concern. In essence, the monitoring service enables a third party to keep an eye on the user's everyday activities in addition to their home area. In light of this, a surveillance system that monitors the home area automatically and uses inexpensive components was designed [7-11]. Crime is defined as any unlawful behavior that violates the law and occasionally endangers the lives of others. Around us, crimes of all kinds, including theft, kidnapping, burglary, and robbery, are always happening. This paper proposes a surveillance system based on computer vision to identify burglary events. Theft means breaking into a specific building or place and stealing items.

In this paper, we propose a surveillance system with a facial recognition that can be used in home and office environments using Histogram of Oriented Gradients (HOG) computer vision algorithm [12-15]. The main system receives camera data in real time, processes image data, and recognizes faces. First, each and every frame will be taken by the camera and forwarded to the central processing unit. A few computer vision algorithms then analyze the frame in order to detect faces and motion. Usually, surveillance cameras are used to monitor a specific area in real time. The motion detection algorithm is used to check whether there is a moving object in the real-time video being monitored. In contrast, the face detection algorithm ascertains whether a human being is present within the residential area. The facial recognition algorithm will begin to function if a face is found. Determining if a face belongs to a stranger or one of the locals is the goal of the face recognition algorithm. In the proposed system, when a stranger is detected while observing the surveillance area, a message is automatically sent to the preset administrator's mobile phone. Users can monitor the contents of surveillance cameras in real time using their mobile phones even when outside, and can check immediately even if there is an outside intruder. Additionally, the user permits the mobile application to sound an alarm, alerting the intruder to the situation.

2. RELATED WORKS

The Haar algorithm is a powerful technique for detecting objects in images, mainly widely used in face recognition. This algorithm uses simple square patterns called Haar-like features to analyze specific regions of the image. Haar features are calculated based on the difference in pixel values between white and black squares and capture the points where a particular shape or boundary appears. Each Haar feature is defined by a specific size and location, and all areas of the image are scanned using a pattern of squares of various sizes and locations. The core of the algorithm is integral image technology, which enables very fast calculations. Using integral images, the sum of pixel values within an arbitrary rectangular area can be calculated in constant time, greatly improving computational speed. Haar features are learned through the AdaBoost algorithm, which creates a strong classifier combining multiple features. AdaBoost builds a strong classifier by selecting the most discriminating features among thousands of Haar features [16-18].

These strong classifiers are typically arranged in a cascade structure, where simple, computationally inexpensive classifiers first filter the image quickly, and then increasingly more complex classifiers analyze the remaining regions in detail. Thanks to the cascade structure, the algorithm can detect faces with high accuracy even in real time. The Haar algorithm is robust to lighting changes and various facial expressions, and demonstrates high performance in many real-world applications. In particular, it is widely used in facial recognition, security systems, and surveillance cameras, and is still considered an important foundation for face detection technology. As a result, the Haar algorithm is an essential tool that has been used for a long time due to its efficiency and practicality in the field of object recognition.

The Histogram of Oriented Gradients (HOG) algorithm is a very important technique for object detection in computer vision [12-15]. This algorithm is mainly used to recognize human shapes in images, but also has applications in various object recognition. The basic concept of HOG is to determine the shape of an object by analyzing the local gradient direction of the image. When analyzing an image, HOG first divides the image into small cells. The pixels within each cell calculate the direction and magnitude of the gradient. Gradient calculations are often performed using tools such as Sobel filters. The calculated gradient direction is divided into angular ranges and expressed as a histogram. For example, divide the angle from 0 degrees to 180 degrees into several sections and record the slope size within each section as a bin in the histogram. This process is performed on a cell-by-cell basis, and the histogram of each cell expresses the characteristics of that cell. Afterwards, multiple cells are combined to form a block. A block contains a histogram of several cells, which provides spatial information and contrast (norm).

This block is converted into the image's feature vector, which is used to extract important features from the entire image. Feature vectors are usually combined with a classifier, such as a linear Support Vector Machine (SVM), and used to determine whether an object exists. HOG has robust characteristics against image rotation, lighting changes, and noise, and can recognize objects under a variety of conditions. In particular, this algorithm has the ability to accurately detect human shapes even in high-resolution images or complex scenes. Additionally, HOG has high computational efficiency, making it suitable for real-time object detection. This is especially useful in real-time applications such as autonomous vehicles, security systems, and surveillance cameras. The HOG algorithm is a fundamental step for object recognition, and the extracted feature vectors can be combined with other advanced algorithms to achieve higher accuracy. These features make HOG a powerful tool widely used in computer vision today. HOG plays a key role in object detection in various fields, and when used with machine learning models, it makes a significant contribution to image analysis.

3. Surveillance system with Face Recognition

This chapter explains the monitoring system proposed in the paper. In particular, the overall structure and functional flowchart of the system is explained in detail.

3.1. *Proposed system architecture*

In this study, we propose a surveillance system that can detect faces using a computer vision algorithm. To verify the proposed system, it was produced using actual hardware and a program was produced using a vision algorithm. The Python language was used to implement the vision algorithm and process real-time images. An Android application using Java code makes up the second subsystem. The local HTTP web server and TCP handshake protocol link the two subsystems. Figures 1 show system process and functions.

The proposed system receives real-time video data from surveillance cameras and recognizes faces. In the actual internal processing system, images received in real time are image processed through a computer vision algorithm and the user is notified of the results. In particular, the system has five major processing functions. First is the alarm function. This is a function that intuitively notifies you with a sound or LED when a specific problem occurs in the system. Second is the real-time video transmission function. Real-time video data transmitted through surveillance cameras is used for image processing, but can also be transmitted to the user for real-time monitoring outside. The third is the motion recognition function. This function notifies you when motion is detected when the system is set to surveillance mode. The fourth is the face recognition function. This function detects outsiders other than pre-registered users. When an outsider is detected, the user can be notified in real time. The last function is real-time message notification function to mobile phone. This function notifies the user in real time when motion is detected in the surveillance system or an outsider is recognized.

Figure 2 specifically explains the system process described above. Surveillance cameras collect real-time video data from the surveillance area and transmit it to a processing system. The video data collected at this time is sometimes converted into still images for image processing. The processing system processes five functions provided by the overall system. In particular, the HOG computer vision algorithm is used for gesture recognition and face recognition. It also has the ability to send messages to mobile phones through a developed program. On mobile phones, you can receive messages from the system and monitor video in real time.

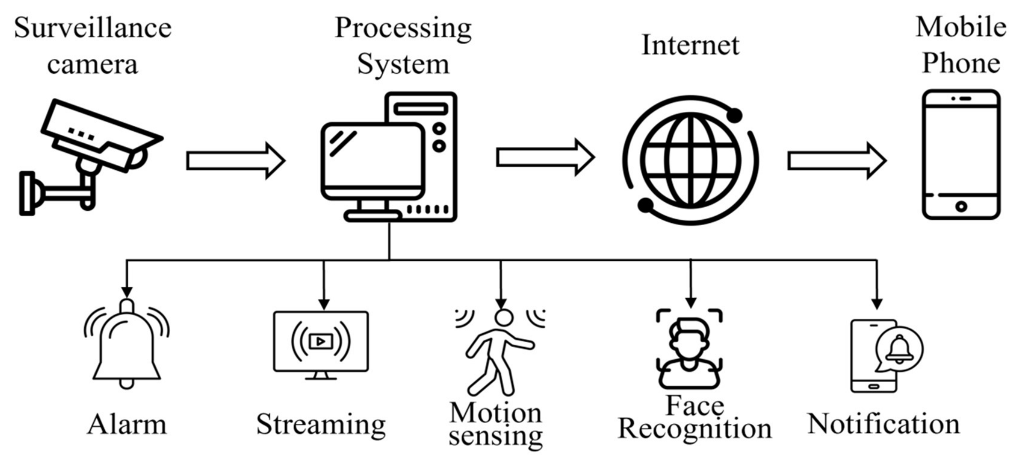

Figure 1. *System Process and Functions*

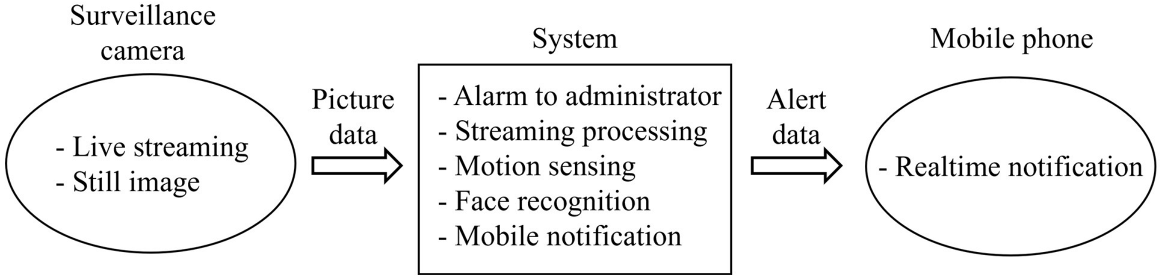

Figure 2. *System architecture*

Figure 3 shows data processing in the system. It is divided into three stages: surveillance cameras, processing systems, and users. Real-time data is transmitted from surveillance cameras to a processing system. At this time, video data can be transmitted wirelessly and wired. The transmitted data is used for initial motion recognition. Verify that the surveillance system is activated and that motion is detected in the surveillance area. At this time, if motion is not detected, new data is continuously received to detect motion. When motion is detected, face recognition occurs, and two cases can occur. If the user is registered in advance, motion detection is repeated again without taking any other action. On the other hand, if the user is not registered in advance, a notification is sent to the user via mobile phone.

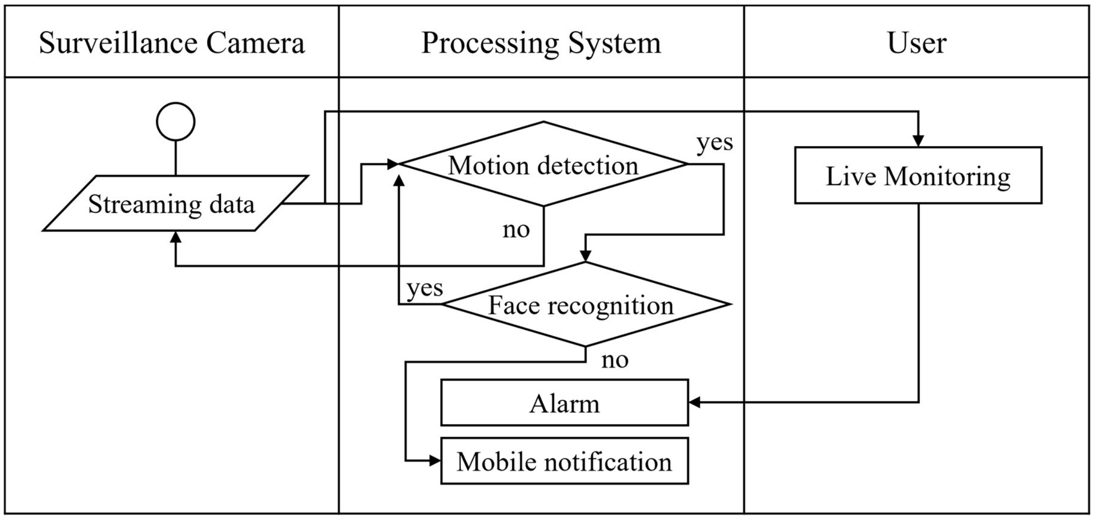

Figure 3. *System flow diagram*

3.2. *Functions for system*

To maximize the system's usability and functionality, the user must configure the hardware and system before utilizing the application. It is advised that the camera be installed at the front door of the residence. For improved face recognition, it's also a good idea to position the camera such that it can catch a person's whole face. The functionality and system interaction procedure are explained in the ensuing subsections.

3.2.1. Image Processing

Initially, the system gathers camera images and uses them to detect movements. The proposed system uses the Open Computer Vision (OpenCV) open source library to process real-time image data. Image processing is performed by dividing real-time video into image frames through the library. The processing system selects one image from the image frame and performs motion detection using the HOG algorithm. Figure 4 shows the image processing steps for busbar detection. Extract the background picture excluding the person from the image frame processed from the video. Afterwards, the difference value between the two images is calculated, and if the value exceeds the threshold, it indicates that there is motion. If it is determined that there is motion, the person's face image is transmitted to the face recognition stage through the foreground mask value.

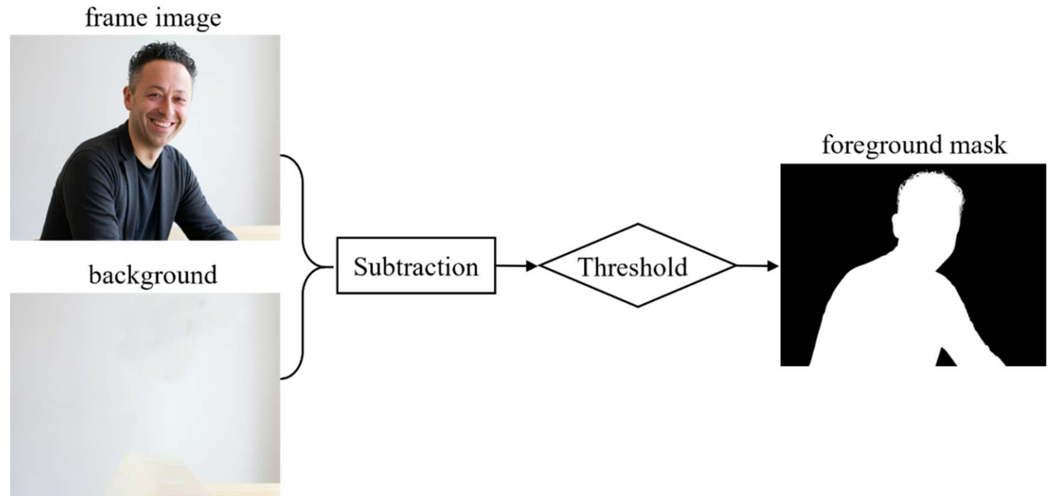

Figure 4. *Motion detection with image processing*

3.2.2. Face Detection

Face detection is performed continuously from the previous step, the motion recognition step. In other words, when motion is detected in the camera image, the image is passed to the face recognition stage. In the face detection stage, one image is extracted from the video and the task is performed. This is to distinguish between pre-registered user recognition, which is facial identification. Such face recognition only works when motion is detected and a face image is transmitted. In face recognition, several tasks such as face discovery, face projection, and face pose are performed simultaneously.

In this system, Histogram of Oriented Gradients (HOG), which is widely used in computer vision, was applied for face recognition[12-15]. This algorithm can be applied in a variety of environments and can be used in bright environments such as offices as well as dark environments such as factories. HOG is robust to rotation and illumination changes and has high computational efficiency, making it suitable for real-time object detection. The first step in HOG is to prepare the still image by converting it to grayscale. This is a technology that identifies the shape of an object by analyzing the local tilt direction of the image. Divide the image into cells, create a histogram of the gradient direction calculated in each cell, and convert it to a feature vector. The image is divided into square sections, each measuring 16 x 16 pixels, to prevent HOG from producing overly detailed results. These vectors are usually combined with a classifier such as SVM and used for object detection. The system then calculates the number of gradient points in each cardinal direction and displays them in the corresponding square using the strongest arrow direction. To confirm that the object found in the motion-detected image is human face, encoding is performed in HOG format. As shown in Figure 5, comparison is performed with the previously created HOG image. The images are compared, and if they are the same, they are identified as the same face.

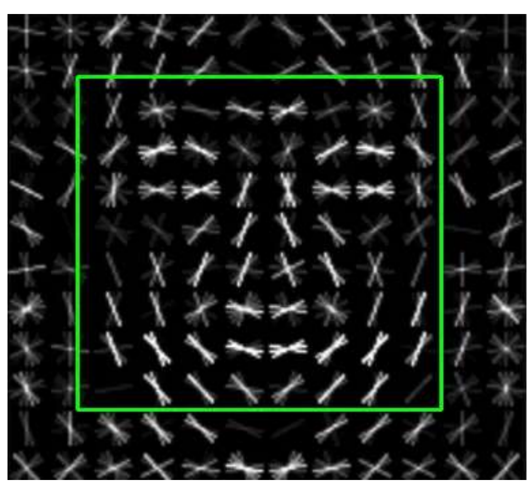

Figure 5. *Face detection using HOG algorithm*

The following step in the process deals with posing and projecting faces, which deals with faces that are angled differently or facing different directions. The facial landmark estimation algorithm is a machine learning technique used to finish the assignment. The system proposed in the paper extracts and distinguishes 68 landmarks from a human face image, as shown in Figure 6. Faces can be found within an image using recognized facial landmarks. Additionally, this information can be used to perform image processing operations such as resizing or rotating the image. And the image of the face can be moved to a specific area.

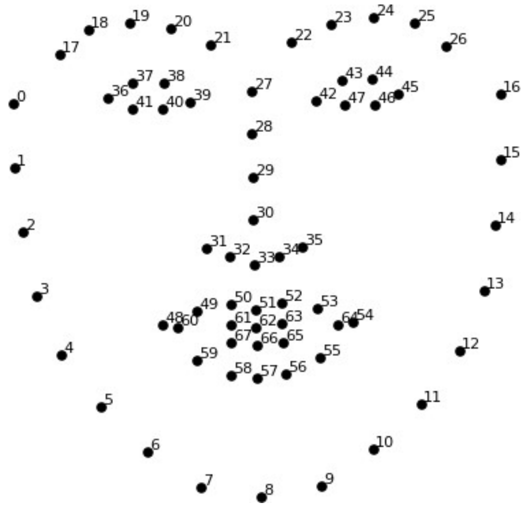

Figure 6. *Landmarks of human face image*

3.2.3. Face Recognition

After going through the previous step of face recognition, the face is identified and its features are created as an image. In the face recognition stage, it is determined whether the transmitted image of the person is an external person rather than a person registered in advance. Existing face recognition methods have the disadvantage of taking a lot of time to compare faces with registered and unidentified people. The paper proposes a new face recognition method to solve the problem of time consumption for comparison in face recognition. The new face recognition method consists of face encoding, feature extraction, and face classification.

Deep learning is used in the feature extraction and face encoding stage to identify the critical facial features that need to be assessed in order to perform face recognition. In order to train the deep convolutions neural network, three pictures of well-known individuals are loaded: one of their faces, one of their faces again, and one of their faces completely different. In this study, the neural network is trained through repeated iterations to accurately generate 128 landmark for each individual. In the development for the experiment, we used the Openface model developed at Carnegie Mellon University to reduce the time. This is because it takes too much time to create a model by directly collecting and learning face images. The system was able to quickly extract 128 landmarks from face images using the model [19][20]. Figure shows extracted landmark values from face image.

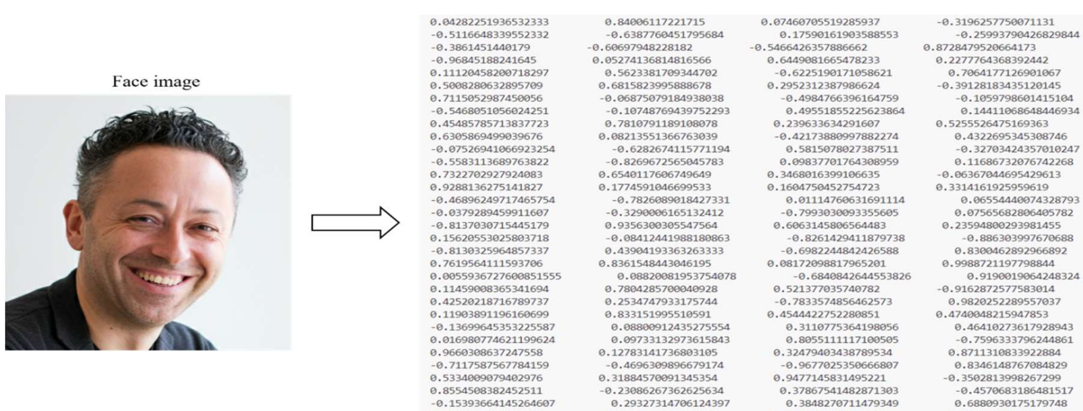

Figure 7. *Extracting landmark from face image*

After 128 landmark points are extracted from the face image, the last stage of categorization is completed in the procedure. Our method swiftly compares previously registered face pictures with newly detected faces using the Support Vector Machine (SVM) classifier. Using SVM, 128 landmarks in newly discovered face photos were compared with existing photos and classified. The result of this comparison served as the basis for categorization.

3.2.4. Notification

Upon detecting a stranger through facial recognition, the system will promptly notify the mobile application. SMS is usually used to send messages to the user's mobile phone, but was not used due to cost. Instead, we used Google's Firebase Cloud Messaging (FCM) API to send messages for free while using an Android phone that is easy to develop. Using FCM, you can send push messages without being dependent on the platform. In general, to receive push messages in real time via a server, the user must always be connected to the server, which greatly wastes the battery and network resources of the user's device. By placing a cloud messaging server in the middle, users can send and receive messages in real time with low battery and network usage. For the above reasons, most application services transmit messages to users in real time via cloud messaging servers.

3.2.5. Alarm

Various methods were considered to provide an alarm function in the system, and the method actually chosen was to use a TCP-based program. To start the alert phase, a protocol to transmit a specific message was selected. When a user engages the alarm trigger feature on their mobile device, the application notifies the system. This particular message is what will set off the alarm on the system. For the purpose of distinguishing amongst function triggers, every message transmitted over the TCP connection is configured differently.

3.2.6. Video Streaming

For video streaming, real-time video from surveillance cameras was converted into a series of still photos using Motion JPEG (MJPEG) technology. The final stage is using the Flask Python web framework API to save the compressed video on a local web server, enabling the web server to broadcast the camera's view. In essence, the user has two options for seeing live streaming: either they can view it through a mobile application or they may view it directly from the web server by first having it stream to the web server. The mobile application might be able to convert JPEG images into videos because it was designed with a MJPEG viewer. Lastly, the mobile phone application downloads the video stream created in MJPEG from the server and plays it.

4. RESULTS AND DISCUSSION

An experiment was performed to verify the proposed system. Through experiments, the performance, scalability, and stability of the system were confirmed. Additionally, the system was able to be improved by collecting problems discovered in the experiment. As shown in Table 1, the system conducted a User Acceptance Test (UAT) to confirm that its features were functioning as intended and that it was meeting the requirements and specifications.

Table 1: *Initial Set of features used for the experimentation*

No.	Step	Expected behaviorz	Result
S01	Server: starting	Server is up and running, waiting for a client	Pass
S11	Client: connect to the server	Ready to pass data, the client is linked to the server	Pass
S12	Server: receive data from client	The server performed a certain task after receiving the message from the client	Pass
S20	surveillance camera: monitoring	The camera's view is retrieved and sent through to detect motion	Pass
S21	Motion detection	The threshold view is visible and capable of identifying motion inside the picture	Pass
S22	Face detection	Determine the face area from the motion detection step. If a face is found, it is marked with a colored rectangle in the image.	Pass
S23	Face recognition	Sort as an unauthorized person or user	Pass
S24	Message notification	The mobile device received a push notification	Pass
S25	Alarm to the mobile phone	Sound system with led display	Pass
S30	Live streaming: starting	View from the camera as seen in the mobile application	Pass
S31	Live streaming: stop	On the mobile application, the camera view is disabled	Pass

5. CONCLUSION

In summary, the suggested system has succeeded in improving security cameras through the use of computer vision algorithms. The system makes use of a number of approaches, including facial recognition algorithms, motion detection, and face detection. First, the system uses the background subtraction approach in between consecutive video frames to detect motion. Second, the motion-capable frames will be routed to the face identification phase for the purpose of identifying the face region. The HOG algorithm was used for face detection and face recognition. Third, face recognition was applied only to the face area to increase accuracy. This article recognizes faces using a deep convolutional neural network architecture.

5. REFERENCES

1. Qaraqe M, Elzein A, Basaran E, Yang Y, Varghese EB, Costandi W, Rizk J. Alam N. PublicVision: A Secure Smart Surveillance System for Crowd Behavior Recognition. IEEE Access. 2024 Feb;12:26474-26491. https://doi.org/10.1109/ACCESS.2024.3366693

2. Kong M, Guo Y, Alkhazragi Om, Sait M, Kang CH, Ng TK, Ooi BS. Real-Time Optical-Wireless Video Surveillance System for High Visual-Fidelity Underwater Monitoring. IEEE Photonics Journal. 2022 Apr;14(2):1-9. https://doi.org/10.1109/JPHOT.2022.3147844

3. Fan CT, Wang YK, Huang CR. Heterogeneous Information Fusion and Visualization for a Large-Scale Intelligent Video Surveillance System, IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2017 Apr;593-604. https://doi.org/10.1109/TSMC.2016.2531671

4. Sun H, Shi W, Liang X, Yu Y. VU: Edge Computing-Enabled Video Usefulness Detection and its Application in Large-Scale Video Surveillance Systems. IEEE Internet of Things Journal. 2020 Feb;7(2):800-817. https://doi.org/10.1109/JIOT.2019.2936504

5. Siddique A, Medeiros H. Tracking Passengers and Baggage Items Using Multiple Overhead Cameras at Security Checkpoints. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2023 June;53(6):3298-3310. https://doi.org/10.1109/TSMC.2022.3225252

6. Liu Z, Lin Y, Liu T, Du B. Reliable Cross-Camera Learning in Random Camera Person Re-Identification. IEEE Transactions on Circuits and Systems for Video Technology. 2024 June;34(6):4556-4567. https://doi.org/10.1109/TCSVT.2023.3341877

7. Suresh MSS, Menon V. An Efficient Graph Based Approach for Reducing Coverage Loss From Failed Cameras of a Surveillance Network. IEEE Sensors Journal. 2022 Apr;22(8):8155-8163. https://doi.org/10.1109/JSEN.2022.3157819

8. Dai Y, Wen C, Wu H, Guo Y, Chen L, Wang C. Indoor 3D Human Trajectory Reconstruction Using Surveillance Camera Videos and Point Clouds. IEEE Transactions on Circuits and Systems for Video Technology. 2022 Apr;32(4):2482-2495. https://doi.org/10.1109/TCSVT.2021.3081591

9. Yan M, Xiong Y, She J. Memory Clustering Autoencoder Method for Human Action Anomaly Detection on Surveillance Camera Video. IEEE Sensors Journal. 2023 Sept;23(18):20715-20728. https://doi.org/10.1109/JSEN.2023.3239219

10. Park JK, Park EY. Implementation of a Remote Healthcare Monitoring System Using Li-Fi. The Journal of Innovation Industry Technology. 2023 Sept;1(2):57-65. https://doi.org/10.60032/JIIT.2023.1.2.57

11. Julia B, Marcel W, Michael R, Michael W, Florian S. Exploring End User Programming Needs in Home Automation. ACM Transactions on Computer-Human Interaction. 2017 Apr;24(2):1073-0516. https://doi.org/10.1145/3057858

12. Zhang X, Zhang L, Lou X. A Raw Image-Based End-to-End Object Detection Accelerator Using HOG Features. IEEE Transactions on Circuits and Systems I: Regular Papers. 2022 Jan;69(1):322-333. https://doi.org/10.1109/TCSI.2021.3098053

13. Alsuhimat FM and Mohamad FS. A Hybrid Method of Feature Extraction for Signatures Verification Using CNN and HOG a Multi-Classification Approach. IEEE Access. 2023 Mar;11:21873-21882. https://doi.org/10.1109/ACCESS.2023.3252022

14. Zhang Z, Zou C, Han P, Lu X. A Runway Detection Method Based on Classification Using Optimized Polarimetric Features and HOG Features for PolSAR Images. IEEE Access. 2020 Mar;49160-49168. https://doi.org/10.1109/ACCESS.2020.2979737

15. Tang Z, Quan D, Wang X, Jin N, Zhang D. Radar Signal Recognition Based on Dual-Channel Model

With HOG Feature Extraction. IEEE Journal on Miniaturization for Air and Space Systems. 2023 Dec;4(4):358-367. https://doi.org/10.1109/JMASS.2023.3299159

16. Brégains RMJ, García-Naya JA, Castedo L. Time-Modulated Arrays With Haar Wavelets. IEEE Antennas and Wireless Propagation Letters. 2020 Nov;19(11):1862-1866. https://doi.org/10.1109/LAWP.2020.2986832
17. Ganguly P, Chattopadhyay S, Datta T. Haar-Wavelet-Based Statistical Scanning of Turn Fault in Vehicular Starter Motor. IEEE Sensors Letters. 2022 Apr;6(4):1-4. https://doi.org/10.1109/LSENS.2022.3158608
18. Moshtaghpour A, Bioucas-Dias JM, Jacques L, Close Encounters of the Binary Kind: Signal Reconstruction Guarantees for Compressive Hadamard Sampling With Haar Wavelet Basis. IEEE Transactions on Information Theory. 2020 Nov;66(11):7253-7273. https://doi.org/10.1109/TIT.2020.2992852
19. OpenFace [Internet]. Brandon Amos at Carnegie Mellon University; 2016 [updated 2016 September 15; cited 2024 Aug 12]. Available from: https://cmusatyalab.github.io/openface/ (website)
20. Sonawane, A., Shekhar, A., Murab, S. A., Pansare, R. K., Satonkar, V. H. & Jha V. K. S. (2024). The Role of Artificial Intelligence in Streamlining University Library Operations. *Library Progress International*, 44(1), 51-66