Iot Based Smart Home Security System with Face Recognition and Weapon Detection Using Computer Vision

M. Nandhini, M.Mohamed Rabik, Kiran Kumar, Ashish Brahma

ABSTRACT: Today home automation systems are popular in households. The control of electric fixtures like fans and lights is possible with the help of Internet of Things (IOT). The problem arises due to intrusion of burglars. The security of such systems has been done using computer vision and IOT. Here we aim to enhance this system by use of image processing for object detection. The system uses cameras at the door for face recognition as access control. Also, vibration and door magnet sensors are installed at the entry points to detect when the burglar tries to barge inside. PIR sensors are employed to detect human presence. A vibration sensor is also used to give alert if any shock nearby is detected. The system allows entry only if authorized person like owner or person registered on the database arrives. The person may be identified through valid proof of identity. It sends a message to the owner in case it doesn't recognize the person within 20 seconds and the owner can monitor the activities via live feed from the camera. All sensor signals are checked and status of the system is updated continuously. In case the burglar tries to break inside, siren is activated and alert messages are redirected to the owner and the police.

Keywords: Home Automation, IOT, Security system, Face recognition, Weapon detection, Vision system, Deep Learning

I. INTRODUCTION

In this era, technology is constantly changing the world. Human brain has been studied extensively in the previous century. Scientists were able to mimic the brain function using complex mathematical models. However, they could not be applied in real life until recently, due to lack of good computing systems. With development of advanced mathematical models, now a computer is able to learn to distinguish between images in real-time. The mystery of why computers could communicate but couldn't recognize a subject in a given image was long unsolved until the introduction of Deep Neural Networks which lead to a new branch of study called Deep Learning. Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." *Neural computation* 18.7 (2006): 1527-1554 was the first paper which lead to the shift from shallow to deep approach towards Neural Networks.

This approach uses various matrix operations in sequence to extract more and more useful features from a given image based on the intuition of how the human brain is able to segregate parts of an image and find useful relations among them. Home security is one of the most significant aspects of our lives. It has been estimated that home security system market is expected to become worth 74.75 USD by 2023 according to a report "Home Security System Marketby Home Type (Independent Homes, Apartments), System Type (Professionally Installed & Monitored, Self-Installed & Professionally Monitored, Do-It-Yourself), Offering (Products, Services), and Geography - Global Forecast to2023". As the homes are getting smarter due to the increase in home automation, the security risks possessed by them become a major concern for businesses and consumers alike. The Smart City Mission envisioned by Ministry of Urban Development (M.o.U.D.) for the term 2015 to 2020 comprises of enabling cities with high quality infrastructure for driving economic growth. This means that internet access will be extended to each corner of the city. The modern homes demand for smart control of fans, lights, ACs, doors, kitchen appliances, etc. In order to meet these needs, each device has to be connected to internet and be remotely accessible by means of a wireless sensor network at any place. Home automation also deals with energy efficiency which can help save on costs. Sensors like thermostats are used to control ambient conditions by using a threshold to govern the indoor conditions. Major brands in home security market have implemented multifactor authentication using biometrics. Video analytics is a major component of most systems where the data generated by cameras is processed for security purposes. The recent trends in home market security indicate that focus is now shifting to providing large scale solutions. The number of devices that can be connected are increasing multi fold. The systems are able monitor indoors and give high quality surveillance feed. They provide protection to the home owners by monitoring levels of humidity, temperature, presence of toxic gases like carbon monoxide, etc.Hence, with the use of improved data driven algorithms, the security systems can be can be enhanced and made even smarter than existing systems in future. For an average household owner, there is no means of monitoring the indoor activities remotely. Most houses don't have intrusion detection systems. Thus cases of theft have increased day by day. Surveillance systems are largely used by offices of banks, government organizations, educational institutions and product based industries.

Revised Manuscript Received on November 01, 2020.

M. Nandhini, Department of Mechatronics Engineering,SRM Institute of Science and Technology, Kattankulathur, Chennai,India.

M.Mohamed Rabik, Department of Mechatronics Engineering,SRM Institute of Science and Technology, Kattankulathur, Chennai,India.

Kiran Kumar, Department of Mechatronics Engineering,SRM Institute of Science and Technology, Kattankulathur, Chennai,India.

Ashish Brahma, Department of Mechatronics Engineering,SRM Institute of Science and Technology, Kattankulathur, Chennai,India.

The existing systems are capable of delivering good performance at much higher costs. There is a scope of reducing costs in terms of both capital and computational performance. Presently, systems are able to do simple object detection for only surveillance and no market solution is offering face access control in home space.In the recent years, more number of solutions relied on cloud services than on edge computing platforms which have seen their entry only recently. Cloud services are unable to deliver low latency. Most systems which are able to offer good speed and accuracy have never been employed in home security space. The most sophisticated large scale surveillance systems to do recognition are employed to monitor road traffic and public spaces. They are capable of whitelisting members of a family. Hence, they have good potential as a component of home security system.Like any other software, AI powered systems are prone to attacks by using AI powered attacks. Images looking very similar to each other but having differences in pixels can be used to create GAN based attacks. This could cause even a high accuracy system to misclassify an otherwise authorized person. This can be tackled by performing adversarial training on the algorithm using hard negative examples.The security of smart homes can be compromised very easily by means of D.D.o.S, P.D.o.S and device hijacking attacks. The identity theft of a person allows an intruder to bypass a system. This can be solved by multi factor authentication. The system must be accessible through physical means only. The tampering of the system can be detected.If the above challenges are not addressed then, there could be catastrophic consequences including threat to lives of the inhabitants. The existing systems offer multifactor authentication via biometrics which means more number of parameters for access. To simplify this, we introduce a single factor of authentication using face recognition. Security systems have provided protection against camera tampering, but it is a late response to a potential intruder. This can be improved using weapon detection along with it.The primary idea was to develop a system with state of the art algorithm on a low end embedded device. The concept can then be used to make efficient security cameras and thereby reducing costs without compromising on accuracy. The need of a system arises as the video data captured can be used for analytics to give details of activities in a timely manner.Most state of the art algorithms have a bottleneck when processing in real time without dedicated hardware. Hence, a need for the system to achieve a reliable result even without high end hardware appeared. The instrumentation of the sensors has to be done so that the house is secured from all sides and not just a single point of entry. This requires the system to be able to detect human presence, if any shock (vibration) is there, when the door or window is moved and an alert signal in case of any emergency.

II. EXPERIMENTAL SETUP

2.1 Overview and conceptualization

After a pilot study conducted on “AI and IOT in Home Security”, it was concluded that the system has to be designed such that it can be deployed in a small form factor such as a single-board computer. Performance and cost of the system usually present themselves as trade-offs. To combat this, we propose a low-end system which can perform surveillance as well as image processing effectively. An embedded system is a resource constrained environment where off-the-shelf algorithms don’t work perfectly. Keeping this in mind, the hardware was chosen such that it met performance requirements and a customized algorithm was developed for the same.The access control methods used previously have been through means like smart cards for residential purposes. Later, sophisticated algorithms were used for biometrics based access control which include fingerprint, retina scanners, etc. However, with advancement in research towards face recognition, the access control methods witnessed the rise of its use in everyday life – the best example being in case of smartphones. It was less secure in its initial phase and is under constant improvement since its introduction into the market.The sensors which were generally used by multinational corporate firms for employees for attendance have inspired ideas for access control solutions. Industries use video analytics for monitoring activities inside the plants. Gradually, these technologies were introduced in homes to make them ‘smart’. Security of smart homes requires design of intrusion detection systems. This may not include physical barriers or traps which are beyond the scope of cost effective solutions. The other sensors which are incorporated into home automation systems include those for illumination, humidity, temperature, fire, seismic activity, etc.The model of a smart home was simplified to include sensors which are part of intrusion detection system only. The proposed system was designed to accomplish two tasks – intrusion detection and access control. Most systems don’t include a method to deal with unforeseen situations like camera tampering. This can be resolved by detecting if wire was cut using the circuit break detection module. An experimental and distinguishing feature of using camera to detect weapons is also proposed to provide alerts about suspicious activities.The proposed system in its current state also incorporates a standard security camera in order to record activities throughout the day which is a common component in most setups.

2.2 System Design

2.2.1 Sensor Module

Spread all around the peripherals and entry points of the house, the sensor module is able to record signals whenever the sensors are triggered. The main door is equipped with door magnet sensor. When, in absence of the owner, the door is opened/forcibly broken, alarms will start ringing in the house and the owner would be alerted with specific notifications. If the intruder tries to force open the door, the vibration sensors equipped on the door will detect the same. Vibration sensor is also integrated with the window panes, a common entry point for burglars. The grill of the window is covered by a wire running through all the parts of the grill. If the burglar tries to cut/weld the grill, the wire will be cut and in turn, a signal will be generated.

A PIR sensor which will be mounted either inside or outside of the house, will detect human presence where there should normally be none. Also present is an emergency panic switch which would be used in emergency situations, either regarding possible burglary while the owner is still inside. Another important aspect for using a panic switch is for senior citizens who are inside the house alone and in need of medical attention.

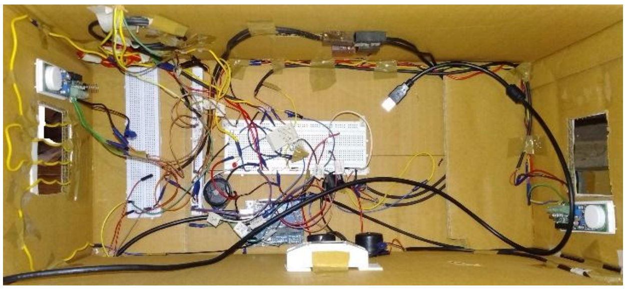


Figure 1 Sensor Module

2.2.2 Camera module

The camera module consists of one primary camera and one secondary camera. The primary camera is Pi Cam, which has been equipped on Raspberry Pi model 3B. This camera is used for face detection as well as recognition. The algorithms are implemented on the Raspberry Pi 3 Model B board.The secondary camera, on the other hand has its own recording and monitoring system. It is a Hikvision network camera equipped with a DVR having the capacity of 1TB HDD. This network camera captures and sends pictures/videos as notifications to the phone whenever any activity is detected.

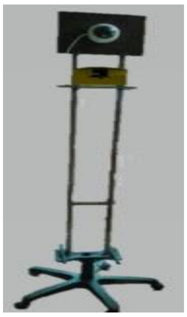


Figure 2 Camera module

2.2.3 Face access control lock

This is door lock which is simulated by a servo motor to which the miniature door model is attached. The servo is connected to Raspberry Pi 3 Model B board via GPIO pins and is fed signal for actuation. Face recognition program runs in loop and as soon as a known face is detected a servo is actuated for access. If the face detected is unknown then, no permission is granted. The person can be distinguished from a thief by detecting if s/he possesses any weapon. For the demonstration purpose, simple tools like hammer, scissors, knife, etc. which are easily available in public datasets are available. This reduces time to development because pre-trained models trained on those datasets are readily available.

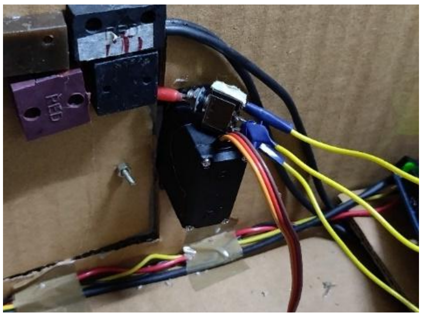
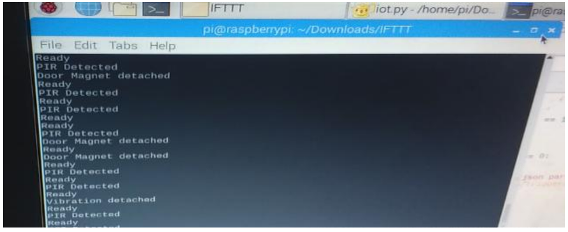


Figure 3 Servo motor for simulating door lock

3.2.4 Notification service

The notifications to the owner can either be on his phone or any remote device. A smartphone/feature phone as the primary device for getting alerts in the form of Notifications, SMS, Calls, etc. using IFTTT app on Android and iOS. Alerts can also be created to notify the police in case of burglary. This module notifies activity with sensors and sends messages when an intruder with weapon is captured by the camera.



```
Ready
PIR Detected
Door Magnet detached
Ready
PIR Detected
Ready
PIR Detected
Ready
Ready
PIR Detected
Door Magnet detached
Ready
Door Magnet detached
Ready
PIR Detected
Ready
PIR Detected
Ready
Vibration detached
Ready
PIR Detected
Ready
```

Figure4 Notification Service using IFTTT app

III. IMAGE PROCESSING ALGORITHMS

The algorithm governing the tasks of face access control and weapon detection were developed. A study on the state-of-the-art methods was carried, giving away clues to the working of these algorithms.Image processing involves getting the image/video data and performing series of numerical computation to give desired output. The entire processing pipeline involves several stages to achieve outputs in various ways such as bounding box detection, ROI landmark detection, labelling of the bounding box, pose estimation, optical flow, gathering information about the activity, tracking a particular object in real time. The image has to be pre-processed before passing it to the algorithm so that inferences are faster. Pre-processing stages include cropping, resizing, blurring, scaling, color transformations, etc. Post-processing is required after obtaining predictions, in order to display the frame in its original form. The goal of the Face Access Control Lock module is to recognize faces in real-time scenarios from the video captured by Raspberry Pi Camera as input and grant permission if the person is authorized. The module's image processing has two sub-modules –

- Face Recognition
- Weapon Detection

The Face Recognition task involves detection of faces in each frame of the video object created for live feed. Detection can be done using Deep Learning based CNN model pre-trained on faces and available as open-source software.

The results of the detection give the location of the faces detected in each frame. For each face detected, a 128-D feature vector is generated by computing the distinguishing distance metrics which can be relative distances between eyes, nose, mouth, chin, jaw line, forehead etc. and also their measurements as obtained using face landmarks. These vectors are then matched with faces which are known with their pre-computed feature vectors. If they match then, the corresponding name stored with the vectors are returned as labels. The end result is a bounding box over the face with name label on top. It is to be noted that the above model used for implementing the above method is a frontal face detector and does not work for other orientations of the faces. Weapon detection is a form of object detection where the key task is to identify the weapons in the frame among the objects captured in the video. Object detection is carried by processing the frames through object detection model which is another Deep Learning based CNN model pre-trained on objects and covers sufficient classes of objects. The objects in the scene have to be detected by segmentation and detections are passed as coordinates for recognition task. The detections of the results are complemented by an additional parameter in the form of a number which is the class id of the object detected. This id is matched with label mapping where the names of the object along with corresponding ids. If the ids match, then the corresponding names are returned as labels. When the label falls under class of weapons, an alert is pushed as notification to the owner. Both the above tasks above require predictions to be made on images. Earlier, primitive methods have been used to make these predictions such as object classification, Image classification and object localization which are relevant even at present. The drawbacks of these were that they were computationally expensive and inaccurate bounding boxes were returned due to mismatch between shape of the box and the object of interest. Recently, methods have been proposed and implemented which can run faster with less number of computations and deliver optimal accuracy. The previous approaches involved cropping of images into equal parts and passing them one at a time to the neural network for detection. However, modern approaches simplify the above step by passing the entire image to the network instead of cropped images. An additional layer in the network is used to compute the edges of the objects all at once. For improving accuracy of the network, 'You Look Only Once' method is used. According to this method, the image is partitioned into multiple grids. The location and classification of the objects is done simultaneously for each grid. These give acceptable results and constitute the group of object classification algorithms came to be known as single-shot detectors. However, they ignore the background information of the image during training. This reduces the accuracy when comparison with two-stage detectors like mask-RCNN networks. The most advanced methods propose tweaking the loss function and not by modifying the network to achieve the desired results.

3.1 Face Detection

Various detection algorithms have been created to detect faces. The first method was developed by Paul Viola and Michael Jones in the early 2000's and widely adopted. It gives the detected eyes and faces as output. More advanced methods are used at present. Some of the methods are as follows-

- Haar-cascades: These methods involve extracting Haar-like features from the given image to convert feature points to feature vectors. One of the popular ways is Histogram of Oriented Gradients (HOG). The method marks arrows on each pixel by comparing the brightness levels of neighbouring pixels on the grayscale image. The collection of such arrows give direction of the pixel intensity as it varies from dark to bright regions. However, the dimensions will be huge for processing. The dimensions can be reduced by consider larger areas of the image instead of going for each pixel.
- Eigen Faces: This method reduces dimensionality of the images by deriving only relevant information from the image. The method uses principle components (PCA) of the image, which are those features with maximum variance, for feature extraction. This reduces the image representation complexity and also saves time and space during computation.
- Fisher Faces: This approach promises to solve disadvantages of the Eigen Faces method. External factors such as illumination affect the PCA, wherein components which may contain discriminative information may also get eliminated. The solution lies in applying LDA (Linear Discriminant Analysis) to reduce variance among the classes instead of maximizing the overall variance. This essentially segregates the same classes from the ones which are different. Hence it is utilized to recognize faces.
- Linear Binary Pattern Histograms: The methods described above work well with lower dimensional data. However, they fail if the amount of data is reduced and noise generated by factors in non-ideal conditions. According to this algorithm, each pixel is given a value of 1 or 0 based on whether it's intensity increases or decreases in comparison to previous pixel's intensity value. The neighbouring pixels create a batch of 3x3 neighbourhood. The pattern of numbers generated in this manner give a fine representation of the image known as LBP codes. These codes can help recognize faces.

More recently, using the above approaches new statistical and probabilistic models have been developed. These approaches are Deep Learning based CNNs which were modified for the purpose of face detection. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015 have proposed a model for detecting faces in images and recognizing them using a unified approach. In this work, Euclidean distances are measured and project same faces near to each other and different faces apart from each other in the Euclidean space.

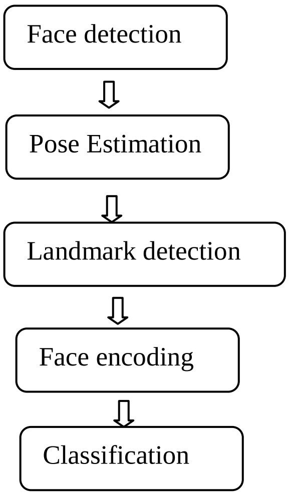


Figure 5 Flowchart depicting face recognition pipeline

The above picture represents the series of steps through which each frame from the video has to undergo during the pipeline. The term ‘pipeline’ refers to the fact that the above steps can’t be executed in parallel but in a sequential manner. However, asynchronous processing of the video thread can boost the speed of the algorithm. The video object can make use of multi-threading for higher number of frames per second. Pose Estimation refers to the process of determining the position and orientation of the head. This is essential as the neural network regards the same person with different head orientations as different. Landmark detection is a pre-processing step for measuring the distance metrics of the face. Kazemi, Vahid, and Josephine Sullivan. ("One millisecond face alignment with an ensemble of regression trees." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014) invented a method to mark landmarks on faces. Their method uses 68 points marked on the faces invariant to the head pose and also resistant to small obstacles on the faces. A dataset of authorized people, with 10 to 20 images per person is used to generate encodings i.e. 128 dimensional vectors for each person. These encodings are later used for comparison with those of faces captured through live feed. Classification involves using a linear classifier such as an SVM (Support Vector Machine) to make predictions when the given face is represented as feature vector. This is similar to one vs all classification as in case of k-means clustering. The accuracy of the face detection model is very high as it was trained on millions of images of faces. Hence, even when very few images of a person are presented it can learn the embedding very quickly. Thus it can classify a person for whom it was never trained before, by just seeing very few examples of the face and the corresponding embedding.

3.2 Weapon Detection

Object detection is a commonly used algorithms in many machine vision tasks for both industrial as well as non-industrial purposes alike. Many methods have been introduced for getting faster and accurate outputs. As the amount of data is increasing day by day, researchers have invested time in building large datasets by carefully putting them in under categories and ensuring that they have rich information for training object classifiers. The notion of using a convolution of operations in succession to extract meaningful features from images has been prevalent long back since 1960’s but could not reach its potential due to limitations on computation. With modern machines, a ‘deep’ neural network can be trained in a fraction of the time than what was possible before..

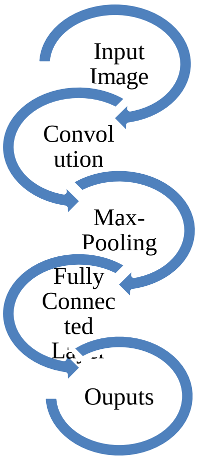


Figure 6Process of Object Detection using Convolution

The object detection algorithms can be used for detecting persons, animals, faces, vehicles and other everyday items. To extract distinguishing features the image is pre-processed and passed through several layers of the deep convolutional neural network. Each layer is a matrix which is the resultant of a mathematical operation performed on the matrix from previous layer. Earlier, a sliding window approach was used to hover over each segment of the image and check if the frame contains an object. Later, the neural networks were made ‘deeper’ by adding layers to do the detection automatically, even when the object is not centered in the sliding window. This is possible with large amount of training examples. The use of convolution in neural networks is crucial for any modern deep learning algorithm, be it face recognition or object detection. The convolution process consists of dividing the entire image into tiny equal parts which can be termed as a tile. Each such tile is traversed through a small neural network and saved into a separate array. If any useful information is available, it is marked for using afterwards. The array is able to map out different parts of the image where useful features are available. The array is high dimensional. So, it can be projected to lower dimensions by a technique called down-sampling. Here, the array is processed in square grids constituting a single batch. The maximum value of those grids is stored in the output array. This operation is known as max-pooling.The matrix which was extracted using max-pooling is an array of numbers and can be unrolled to give a feature vector which has a single column of numbers. This is basically an array which is further passed into another neural network called fully connected layer. The role of this neural network is to make a prediction of whether the given image matches the object in question. The above 3 steps are repeated in various combinations to get highly complex networks. This way helps to learn even more complex features, as the number of convolution layer increases. The goal of the process is to reduce the image into simpler feature vectors which can help in distinguishing the images.

3.3 Internet of Things

Internet of Things can be defined as a network of devices called ‘things’ connected and controlled via the internet. The connection can be local or global. The ‘things’ are mainly analogue data sources. They can be simple devices such as smartphones, smartwatch or even sophisticated such as machines, tools, cars, clothes, people, animals, buildings, etc. Things are connected to data acquisition systems which store the sensor data. These systems transmit the data through Internet Gateways. The processing of this data must be done so that it is ‘cleaned’ and only desired information is kept for records. Finally, the processed data is stored in Data Centre also known as ‘Cloud’. This architecture helps in loading off the processing burden on the things. All the above steps are managed by an Analytics Management Control system. It helps visualize the data being transmitted from each stage.The sensors/actuators can be controlled by using a network known as Wireless Sensor Network (WSN). This type of network is created by allowing each of the things to have an independent wireless adapter. The sensors are connected in mesh like structure such that the information can easily travel in P2P (peer-to-peer) manner. The sensor is then known as sensor node.Sensor nodes are connected to routing nodes which are basically routers. This is analogous to the internet connection supplied by a service provider in an apartment. Each tower can share a common router cable connection line. The information travels through router nodes in a way similar to the routers of each flat, the difference being that in case of the latter the point of origin is same whereas the former is flexible in terms of the path through which the information passes.

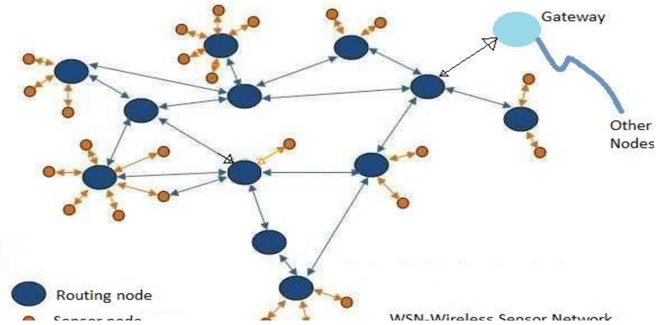


Figure 7Wireless Sensor Network

The above architecture was implemented by configuring Raspberry Pi board as the router node and the sensors connected to it as sensor nodes. No Gateways were configured specifically as a dedicated API was used to program the notification service.IFTTT is an application which helps connecting different apps and services together. It eliminates the need of any other expensive communication devices. The Webhooks service is used in the program to read a sensor value from Raspberry Pi and push a notification using a POST request. Internet of Things can also be implemented for sending images of an intruder captured in the camera’s live feed to WhatsApp by a service called Twilio. This service helps in automating chat messages. It is a convenient way to redirect messages to the owner.The sensor data can also be sent and an event can be triggered. In this system, the event (alert) gets triggered when the sensor value reads a high signal value. The alerts can be triggered to call the police if the person tries to break into the house.The cyber security aspects become pivotal in determining the effectiveness of the system. No system is fail-safe at some point. Hence, to avoid attacks which can be used to spoof the access control and the manipulation of sensor values, multifactor authentication is proposed as a preventive measure.The systems must be provided with anti-DDoS software modules. This however requires years of experience in Cyber Security and is beyond the scope of this project. The modules are expensive and hence were not incorporated for this scaled down model of the system. However, proposed system can incorporate this feature, given that the required infrastructure consisting of hardware and additional costs is in place.

IV. SYSTEM EVALUATION

The primary objectives of the system are:

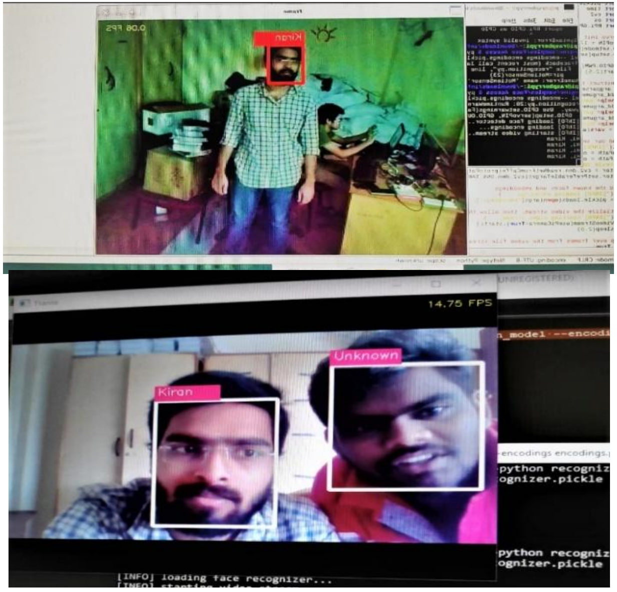
- Perform Face Recognition using higher accuracy models
- Detect weapons in video stream
- Send timely alerts triggered by sensors
- Detect tampering of the system

The evaluation of the system was done by using simple metrics such as speed, latency, accuracy, etc. All the parameters considered depend highly on the methodology followed.

4.1 Testing and Benchmarks

4.1.1 Face recognition model

The Face Recognition model uses deep metric learning present in Dlib C++ Library. The model reached an accuracy of 99.38% on the Labelled Faces in the Wild (LFW) dataset. This network was trained on 3 million images. It uses a ResNet having 29 convolutional layers.



```
[INFO] loading face recognizer...
```

Figure 8 Authorized and unauthorized person detection

The number of frames per second was considered as evaluation metric. The time taken to generate a frame also includes the time required to perform inference on a single frame. The trained model for face recognition is converted to inference model which can be used with the program.

Inference means that the trained model is tested on various test cases and if the performance is satisfactory then, it is ready for deployment. The system's performance was evaluated on 3 different working environments.

Table 1 Face Recognition results using face_recognition by Adam Geitgey

Configuration	Frames per second (inference time inclusive)
Raspberry Pi Model B + Neural Compute Stick 2 (OpenCV 4 + OpenVINO)	0.5
Intel Core i3-4500U CPU	1.03
Nvidia GeForce GTX 1050 Ti GPU	14.75

The above results were obtained as the model was taken directly without any modifications in its architecture or inference model.

The above results indicate that even though we added Neural Compute Stick 2 hardware accelerator to Raspberry Pi, the performance dropped as the forward propagation of the network was done through the ARM processor and not the hardware accelerator. The deep learning model does not perform well on low-end CPUs. The GPU performance is as expected as the forward propagation is much faster in an Nvidia GPU with CUDA support. In order to improve speed another model called Deep Face was adopted.

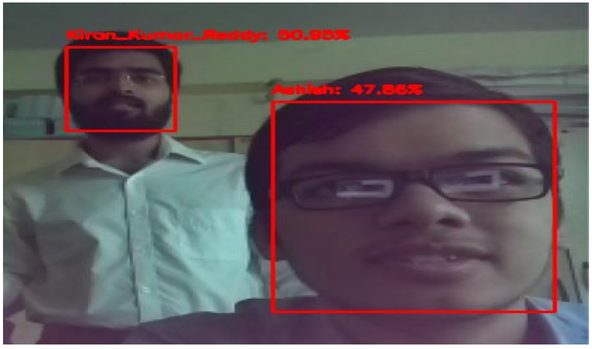


Figure 9 Testing deepface model as an alternative

The model was tested for all 3 configurations. It was observed that it could handle non-frontal faces. It was not a highly accurate model and required retraining. The speed was enhanced over 10 times as both face detection and embedding (face descriptor) models were defined using OpenCV's dnn module. This could easily take advantage of the hardware accelerator. The performance results improved due to which the video frames appeared as real-time.

Table 2Face Recognition results using DeepFace model

Configuration	Frames per second (inference time inclusive)
Raspberry Pi Model B + Neural Compute Stick 2 (OpenCV 4 + OpenVINO)	4.7
Intel Core i3-4500U CPU	3.51
Nvidia GeForce GTX 1050 Ti GPU	21

However, this model had a considerable amount of false positives and false negatives during testing. Hence, the precision and recall values were low for this model. The custom dataset used for generating the face embedding were same. Deep Face model had face embedding trained on an SVM which was used to generate probabilities for a match between faces and name label. This is not effective when compared to the Dlib Library's face descriptor which uses exact face measurements and computes the feature vector using statistical approach instead of a probabilistic model. Neither of the models above fall under 'sweet spot' in speed vs accuracy trade off. The acceptable solution was to put accuracy as a priority. Hence, the face descriptor from Dlib was chosen as the final model to be deployed.

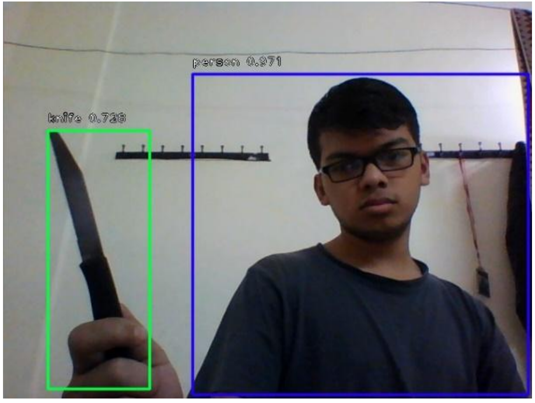
4.1.2 Weapon Detection

Weapon Detection uses the RetinaNet object detection model adapted in Keras library. It is an extension to SSD-ResNet model with an additional convolutional neural network called Feature Pyramid Network as backbone which is used for bounding box regression.The performance was poor for this model. This model architecture being computationally intensive and with a mAP score of 32 on the COCO dataset requires high performance GPUs with lots of memory to give faster throughput results. It is a highly accurate model. However, there were difficulties during classification for various orientation of the same object. This could be improved with retraining.

Table 3 Weapon Detection results using keras-retinanet model

Configuration	Frames per second (inference time inclusive)
Raspberry Pi Model B + Neural Compute Stick 2 (OpenCV 4 + OpenVINO)	No Support
Intel Core i3-4500U CPU	0.07
Nvidia GeForce GTX 1050 Ti GPU	0.24

The model is officially not supported to run Raspberry Pi and hence no results could be obtained for the same. The solution for increasing CPU performance is to use Open VINO library and covert the model into intermediate representation file which can be used across various plugins. The Open VINO library supports high end VPUs for image processing. It is expensive to deploy in a scaled down model. Hence, there were no further experiments conducted in this regard.



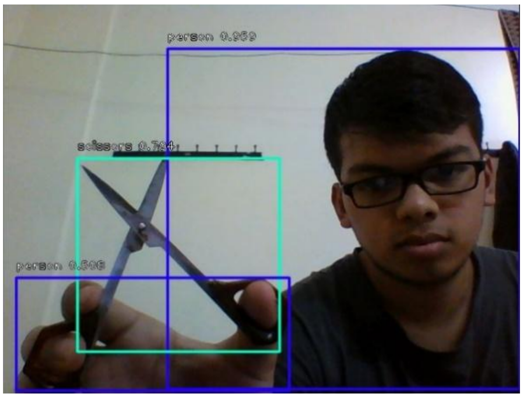
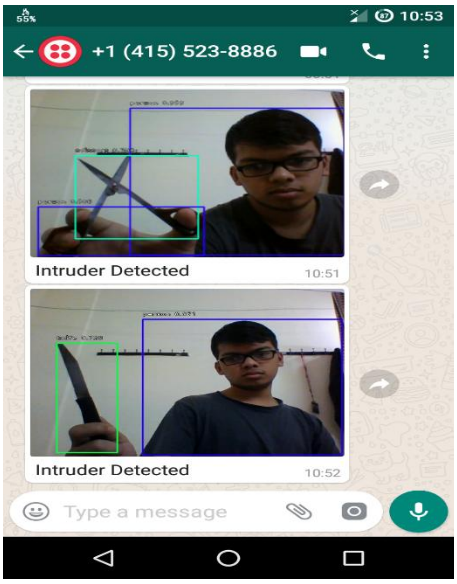


Figure 10 Weapons detected and labelled with their names



+1 (415) 523-8886

Intruder Detected 10:51

Intruder Detected 10:52

Figure 11 Notification alerts received via Twilio API

4.1.3 Sensor Module

There were three sensors namely- PIR sensor, Vibration Sensor, Door Magnet sensor. These were tested rigorously for sending alerts with low latency. For the purpose of demonstration, their range and sensitivity were adjusted. The range was reduced whereas sensitivity was kept high. This allowed for faster testing. The IFTTT app was used to receive notifications in the app itself as it was the most reliable way. It is also configurable for SMS alerts. The alerts were slow during initial development due to the use of Arduino IDE for sending notifications. Later, with the help of Python requests module, the POST method was used and it gave very quick notifications.

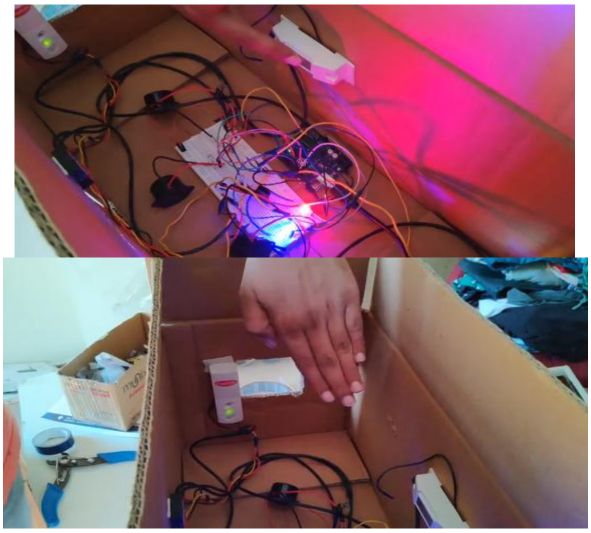


Figure 13 Working of Sensor Module

The system was also able to detect if there is a circuit break using Wire Break Detection which is placed on the grill of the windows. Hence, all the objectives were successfully met by the system.

V. CONCLUSION AND FUTURE SCOPE

The system developed in this work proposes a simplified way to handle access control and intrusion detection in smart homes. The use of face recognition for access control is an efficient way for this environment. The sensors used are capable of providing instant alerts, so that the owner is never left to unforeseen circumstances. This project clearly shows how state-of-the-art methods can be used even with limitations on hardware. The design of the system for achieving the ‘sweet spot’ between speed and performance is an iterative process. It requires further investigation with customized models and training on large datasets. Therefore, the best performance can be achieved by redefining image processing algorithms and not using off-the-shelf models. One of the important highlights of the system is weapon detection using image processing. The primary advantage of it is that it helps avoid expenditure on any physical components, thus reducing costs significantly. However, it is an experimental feature which cannot be incorporated into a single model alongside face recognition. This method promises to improve the detection rates of a potential intruder and improve the overall efficiency of the access control system. The system also implements a fail-safe method by having extra components powered by electric power backup in case there is a circuit failure. The tampering of the circuit is also detectable, which makes it suitable for emergencies. With all the above features, the system is apt for use in modern homes and forms an essential avenue for building smart cities. The system can be further used in other industrial areas such as logistics, schools and ATM facilities for improving security. This will reduce the amount of time taken to identify threats to the buildings where the system is employed.

FUTURE SCOPE

Even though, the final design was found as satisfactory, it was realized there is always scope for improvement. The system is not entirely break-proof in spite of providing fail-safe method. Some of the aspects worth noting for future improvements are as follows-

- Liveness Detection System: During the process of development of Face Recognition system, it was observed that the system can be spoofed by using a photo of the owner for access. Hence, a liveness detection system was incorporated during the testing of the system. The system is able to distinguish faces as valid or invalid based on the whether the captured face is from a phone or the person himself. However, the disadvantage of the system was that it doesn't generalise well on all faces. Actual methods also include detection eyes blinking, motion sensing, how pixels change and 3D depth sensing. Depth sensing required cameras with depth sensors for determining if the face is a 2D image or a 3D solid object. Due to lack of suitable hardware and low recall of the method, it was discarded in the final design of the system. Nonetheless, it forms a crucial part in improving the system.
- Non-Frontal Face Detection: Only front faces were detected in the final model. The alternative model using DeepFace could allow for non-frontal faces as well. The descriptor of the current model can be combined with it to achieve recognition of non-frontal faces with very low latency.
- System Performance: The lower performance of the system can be compensated by using dedicated GPUs present in embedded platforms like Nvidia Jetson boards and Google Coral TPU board.
- Camera Tampering Detection: It can be created by using image processing and predicting by rate at which frames change.
- Indoor Surveillance: Intrusion and access control methods provided above can be also used with indoor surveillance where the cameras are used for monitoring infants in case parents are away and check if pets are safe.
- Fire alarm: The system can be given an extension to alert the owner if a fire was set in the house when they are away. This could help in immediate evacuation.
- Sensor Nodes: In order to make sensors portable and accessible throughout the globe, they can be assigned with individual IPs for their wireless and 4G LTE adapters. This could help in accessing their data globally.
- Drone surveillance: The above system can be incorporated in drones for close quarter surveillance operations and identification of people from database.

REFERENCES

1. Dubal, P., Mahadev, R., Kothawade, S., Dargan, K., & Iyer, R. (2018). Deployment of customized deep learning based video analytics on surveillance cameras. *arXiv preprint arXiv:1805.10604*.
2. Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." *Neural computation* 18.7 (2006): 1527-1554
3. Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
4. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.
5. Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.
6. Mao, J., Lin, Q., & Bian, J. (2018). Application of learning algorithms in smart home IoT system security. *Mathematical Foundations of Computing*, *1*(1), 63-76.
7. Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." *bmvc*. Vol. 1. No. 3. 2015
8. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
9. Wen, Yandong, et al. "A discriminative feature learning approach for deep face recognition." *European conference on computer vision*. Springer, Cham, 2016.
10. Dlib C++ Library by Davis King http://blog.dlib.net/2017/02/high-quality-face-recognition-with-deep.html
11. Evolution of Object Detection and Localization Algorithmshttps://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad
12. Face Recognition API by Adam Geitgey https://github.com/ageitgey/face_recognition
13. fizyr/keras-retinanet - Keras implementation of RetinaNet object detection.https://github.com/fizyr/keras-retinanet
14. Home Security System Marketby
15. Home Type (Independent Homes, Apartments), System Type (Professionally Installed & Monitored, Self-Installed & Professionally Monitored, Do-It-Yourself), Offering (Products, Services), and Geography - Global Forecast to 2023
16. Install OpenVINO Toolkit for Raspbian OS https://docs.openvinotoolkit.org/latest/_docs_install_guides_installing_openvino_raspbian.html
17. Machine Learning is Fun! Part 3: Deep Learning and Convolutional Neural Networks https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721
18. Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78
19. PyImageSearch by Adrain Rosebrock https://www.pyimagesearch.com/