2018 IEEE International Conference on Innovative Research and Development (ICIRD)
11-12 May 2018,Bangkok Thailand

Security using Image Processing and Deep Convolutional Neural Networks

1st Goutham Reddy Kotapalle
Software Engineer
Cisco Systems Inc.
Bengaluru, India
goutamkreddy@gmail.com

2nd Sachin Kotni
Software Engineer
Walmart Labs
Bengaluru, India
kotnisachin1995@gmail.com

***Abstract*—Safety has, for a long time, been one big thing everyone is concerned about. Security breach of private locations has become a threat that everyone intends to eliminate. The traditional security systems trigger alarms when they detect a security breach. However, the usage of image processing coupled with deep learning using convolutional neural networks for image identification and classification helps in identifying a breach in an enhanced fashion thereby increasing security furthermore to a great extent. This is due to its capability to extract complex features from the images using accurate and advanced face and body detection algorithms. The rate at which machine learning, especially deep learning, is transitioning is very high. The use of such technology in taking the existing systems and models to the next level would be a great step towards advancements in every field of science and technology. The same goes with computer vision. These two coupled and brought together to be used in the field of security results in achieving a lot more than what is imagined to be possible and this paper aims to do the same.**

Keywords*—Motion Detection, Image Processing, Neural networks, Open CV, Tensor Flow, and Microcontrollers.

I. Introduction

Technology used in securing highly important places has changed a lot since the last few years and will continue to change in the coming years. Security is very important when it comes down to smart applications. The new and emerging concept of smart security offers a convenient, comfortable, and safe way for securing highly sensitive areas [1]. Security systems used conventionally aim to protect a place from a breach by sending a notification in the form of a triggered alarm at the time of breach. However, the proposed security system offers many more benefits when compared to the conventional systems which are discussed in detail as we go further ahead into the implementation and working of this system.

This paper focuses on how security at locations considered very sensitive and private such as a location where highly valuable or sensitive data is stored can be made much more effective by deploying intelligent systems that are capable of performing with efficiency levels that cannot be achieved by a human or even other traditional security systems. This system comprises of two modules defined at the hardware level which includes a Raspberry Pi Microcontroller with a few sensors connected to it and an Arduino Microcontroller with Global System for Mobile Communications (GSM) and Global Positioning System (GPS) capabilities installed together at the area of deployment. These two modules communicate with each other on the local network and together communicate with the users on the remote public network.

This paper also focuses on how computer vision can be used to detect human activity at the site and the use of deep convolutional neural networks to identify and match an image with a set of people authorized to visit a site. Thus, with a combination of both, the efficiency of the system is increased multifold.

II. Motivation

Technology is evolving a lot and there have been many advancements in the field of Internet of Things (IoT) as well as in the field of security. This system aims at providing enhanced and much stronger security. The use of computer vision and machine learning has helped in further enhancing the system. This has reduced the number of sensors to be installed at the place to be secured and hence allowing for a minimalistic hardware setup. This system is supported by optimized image processing and machine learning algorithms which are placed at two consecutive levels of the system's software architecture where in consecutive frames of a video are taken as input which are then processed to output well formatted data which contains the breach information if there is any.

Machine intelligence is one of the hottest topics in this technology era that gives computers the ability to learn things by taking input from the surroundings and taking actions to maximize the chances of successful output. The process involved is similar to the approach followed in data mining. Data mining extracts data whereas machine learning tries to find patterns in the data. Here, the images of the users that are authorized to access the secured area are considered to be the initial data set used to the train the neural network model used in the identification of a face. The system would then be able to recognize the faces of particular users which will help us take appropriate decisions and in turn make a more effective security system. Consequently, the addition of new users to the system would require the neural network model to be retrained in order for it to achieve the capability of identifying the newly added

978-1-5386-5696-9/18/$31.00 ©2018 IEEE

faces. This sets new values to the parameters used in the neural network model.

The use of computer vision has gained momentum lately and the algorithms in this field are getting better and better with time consequently helping in achieving better results. This paper also focuses on image processing where we intend to detect faces from the frames obtained using the hardware camera modules sending data to the microcontroller modules where the image processing algorithms are present. The algorithms in turn pass the processed feature rich images to the neural network model present at the next layer of processing. The system follows a three tier architecture where each level has its own importance where the first layer would be the hardware receptor layer followed by the image processing layer above which is the neural network layer.

III. Literature Survey

A. Survey of the Existing Models:

There has been a lot of research going in the field of security and Internet of Things especially in the last decade. The technology being used in achieving smart security is very diverse and huge. In some existing systems, whenever a breach is detected using the camera modules incorporated in them, images are captured and a mail is sent to the user, having detailed information about the situation [4]. There have been smart security systems which have been built very cost effectively which could lead to the security being compromised to some extent. The short text message based security systems [2] are based on a technology where a short text message (SMS) can be sent to the system to activate or de-activate a particular service which would in turn reply back with a response in a similar way.

There have been many more enhancements in automated smart security systems with the integration of smart phones which would be used to communicate with the main modules of the system residing at the area to be supervised. There have been integrations done with smart phones such as controlling the security system using voice commands using special application program interfaces [11] or by connecting the smart phone to the security system with the help of Bluetooth [12].

Before the use of smart phones, the legacy systems used the internet along with text messaging technology to monitor and control a system deployed in an area. In such cases, a web server would have to be hosted in their core hardware modules which would have all the sensors connected to them. When the user wants to trigger something or when a security breach occurs, a client side web application would act as universal tool for any kind of interaction with the system behaving like an administrator panel to control and view the status of the entire system through the browser [10].

With current advancements in machine learning and image processing, a lot of simplified models have been developed which are computationally less expensive. With these generalized machine learning and image processing models were developed systems which were capable of recognizing any objects in images by being able to draw fine boundaries between them. Machine learning in such systems made use of convolutional neural networks which are hybrid neural networks highly used in face recognition and object detection [1]. These networks extract complex features from the images and are able to classify and detect objects in the images.

Live streaming is another way to keep track of the activities at the area of deployment but there should be some amount of manual effort put in such cases [8]. But this again is a very costly approach and is accompanied by storing and retrieving large amounts of data thus making such systems not a very ideal solution. For further enhancements there has also been motion detection incorporated into the live streaming approach where each and every frame is compared for any change in pixel values of the frames [9].There has been research done in this area where the video is sent to the cloud and every frame in each second is being checked for new objects which increases the complexity of processing and puts a tremendous amount of pressure on the network bandwidth due to the amount of data sent to the cloud.

IV. Proposed Work

A. Overview of proposed work:

The proposed security system is designed to increase efficiency and accuracy in security by using improvised and advanced object detection and highly efficient convolutional neural network models. The security system contains a microcontroller module which must be placed in the area to be monitored and would act as the core module of the system. This Module would only be used to capture all the data that it senses in the house and uses image processing and object detection algorithms to detect any breach. The detected breach-data is then passed to the cloud for further processing as an input to the machine learning model which then outputs the properties of the image captured based on its previous training. The edge module placed in the area to be monitored contains the user log which is used to register users. Any changes to the system thereafter must be authorized only by the registered set of administrator users. The core module is connected to an external database which saves the timestamp of the breach as well as the breach-data including the video frames containing this information. Before doing so, the video frames with large difference between them are first processed using a first level image processing algorithm.

This data that is stored would contain possible breach information which is further passed as an input for processing to the modified face and body detection algorithm which would identify additional features in the images. The output from this phase of image processing is passed as an input to a previously trained machine learning model which matches the features of the objects detected with the existing database of images of people authorized to enter the secured area. This entire process is initiated by a combination of infrared and thermal cameras capable of capturing high quality video streams in very dark conditions. This camera setup is a part of the core module which constitutes the core hardware of the system configured in the area to be monitored. This data, when received by the database is then forwarded to the client mobile application which displays these details to the end user in the form of an application notification when the user is connected to the internet or as a Multimedia Messaging Service (MMS) message if the user isn't online. Data is further sent out as an SMS and a

978-1-5386-5696-9/18/$31.00 ©2018 IEEE

notifying call to report the user of any breach. This is achieved by using the cellular network module or the GSM module which is attached to another microcontroller sub module that is bundled with the core module and is used accordingly depending on the availability of the user over the internet.

When any authorized personnel are present in the area where the system is deployed, the system understands that the user is in its vicinity using location tracking and gets deactivated by itself and is activated when a change in environment is detected. It must also be noted that this system is very cost effective bearing in mind the design and the functionality that it possesses. Once there is a breach detected by the face and object detection algorithms, the video frames are processed by these algorithms at the core module for detection of any faces and if there is a face detected then the frames are sent to the user along with the face marked with the boundaries using the same image processing algorithms.

The user, on receiving the details, can take appropriate action. Parallelly, an automatic reporting is made to the concerned authorities. The system is deactivated when the user is within the specified geometric coordinate range of the secured area which is calculated using the latitude and longitude values continuously being sent to the system by the client side application on the user's end.

This system's core functionality is based on a computer vision algorithm using an enhanced version of the OpenCV object detection algorithm in the first layer of processing followed by the usage of tensor flow python packages for developing a neural network model in the next layer of processing to detect a possible breach. The algorithm used and the procedure followed have been discussed in detailed in the following sections of the paper. This system also focuses on providing access to additional users upon the authorization by the administrator. If there is any other type of breach other than a security breach such as a fire, then the system notifies the respective departments and authorities of the same and a recording of the activity from which sensor the alarm was triggered would also be sent to them.

V. Implementation And Analysis Of Proposed System

The core module of this smart security system is what is fundamentally used to check the status of the secured area which is done with the help of the camera module which captures a video which is then worked upon frame by frame by the algorithm developed. Whenever the system is active, a continuous frame sequence of the activities at the secured area is recorded and an alarm is triggered according to the situation based on the kind of threat. Here, a modified version of the delta and the threshold functions of OpenCV library's open source algorithms are used to achieve this.

A. *Methodology:*

There are many methods through which the foreground and background can be segmented from which the difference in pixel values is calculated. Why are we focusing so much on foreground and background? Basically when the system is active during the times when there is not much change in the consecutive frames, the video is largely going to be static since there is not much change between any two consecutive frames and thus will result in a mostly constant foreground and background. If there are any major changes between the frames, it would result in a very different foreground in the consecutive frames and a comparatively smaller change in the foreground. But many systems fail in a real world scenario due to shadowing, reflections, lighting and many other reasons that disturb the image quality which are also to be kept in mind while designing the algorithm. One of the most important things to be considered here is being able to make the processing computationally less expensive since the manipulations are being carried out on a microcontroller like raspberry pi. Moving to the next stage of processing in this detection, the server side processing bears scripts which uses Google's open source tensor flow machine learning package which takes a frame as shown in Fig. 1 which is initially uploaded as the input by the raspberry pi client when a change in frames is detected. This image is processed by the tensor flow client libraries and checked for a match with the images of the people who are authorized to enter the secured area. The algorithm sequence that does the first level of processing which is detecting the introduction of new objects in the frame using the previously mentioned modified object detection algorithms.

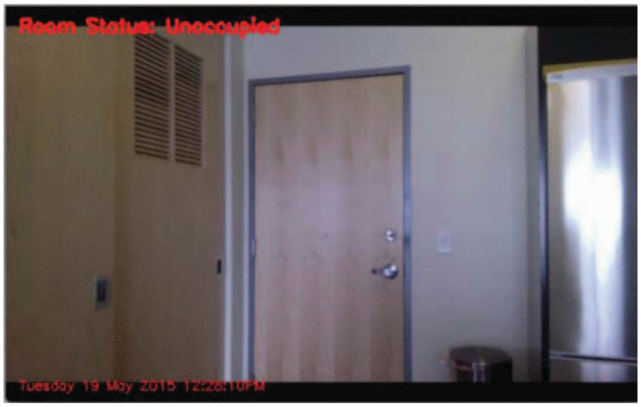

Fig. 1 A frame from the video stream

Once there are two consecutive frames, the processing is going to begin. The images will be converted to a certain size and then change them into grayscale images since RGB frames bear information which is of no interest in motion detection and hence the color is being eliminated and they are converted to grayscale images. Then a Gaussian transformation is applied to the images to smoothen them. As any two consecutive frames will never be the same, there will be some noise in the images due to the inevitable minor background disturbances, so they are most certainly going to have different intensity values. This will help in getting rid of high frequency noise.

Consider two consecutive frames (Fig. 1 and Fig. 2) of the video, the subtraction between the two frames of the video stream is carried out and the absolute value of the difference is calculated. In the next step, the threshold of the difference of the frames is calculated which will help in identifying the pixel values which have significant changes in it. Consider the frame in the video stream as shown in Fig. 2. The marked part in the frame is the change which is detected after processing the part of the frame. The delta of the frames has been carried out on the Gaussian transformed frames so as to eliminate high frequency noise and other noise which could be created due to some

978-1-5386-5696-9/18/$31.00 ©2018 IEEE

natural means and also due to the camera per se since there can be some change in the intensity values of the frames captured by the camera.

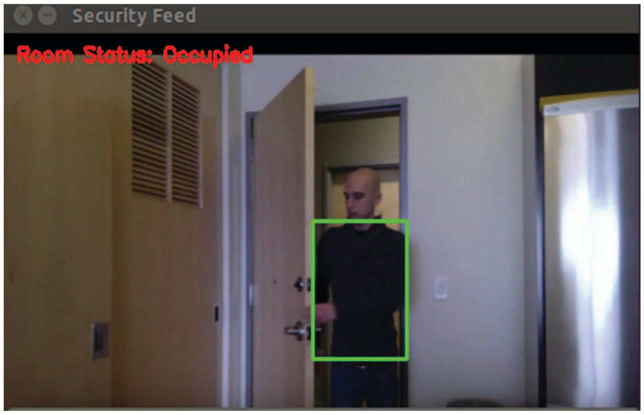

Fig. 2 Another frame from the video stream

This delta, which is shown in Fig. 3, is the absolute difference between the frames. In Gaussian transformation and Gaussian blur, the signal is the image which has to be converted to remove the high frequency noise. It is mainly based on Laplace and Fourier transformations. This is widely used in reducing the noise and also the detail in the image. Gaussian transformation is the same thing as doing the Gaussian blur using the Gaussian function on the image as shown in (1) and (2) on each pixel of the image. Here y is the distance from the origin in vertical axis, x is the distance from the origin in horizontal axis and σ is the standard deviation of the Gaussian distribution.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

In two dimensions, it is the product of two such Gaussians, one in each dimension:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

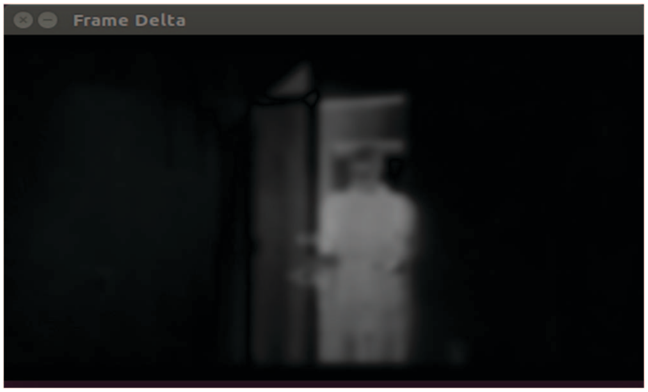

Fig. 3 Delta of two frames

In order to make the above delta significant, a threshold function is applied on it i.e. the pixel values are checked if they are below or above the threshold values and their values are updated if there is a significant change in it so it can easily be noticed and that area is marked in the initial image. Fig. 4 is what is resulted when the threshold function is applied to the delta of the frames shown in Fig. 3.

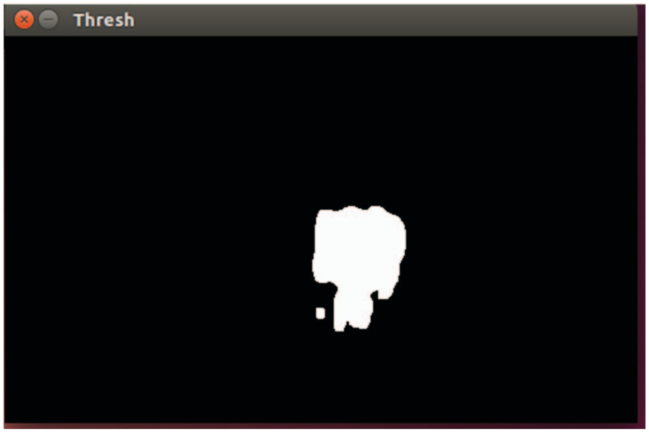

Fig. 4 Threshold applied to the delta frame

This significant change helps in identifying the moving objects in the background and notifying that there is a breach and that the site has been occupied. When there is a change detected, the image will be taken and pushed into a database for future reference which is then fed into an already trained neural network which matches the image with existing photos.

Unlike human brains which easily detect and recognize faces and objects as soon as they are seen, making a system identify an object requires a lot of training. It has been found that deep convolution neural networks can achieve a high performance on hard visual recognition tasks, sometimes exceeding human performance. The image argument is supplied to the model built using the Google's standard open source tensor flow library as a square 299 X 299 RGB image since the model is pre-trained with images of this size. The pixel values don't have to be scaled to be between 0 and 255 because the frames are already converted into that particular format at the object detection phase itself before receiving it at the cloud. The hyper parameters which best fits the model must be set which are used to fine tune our optimization algorithm. These values vary depending on the type of video input and the type of model used and are updated as the model is trained. The graph described by the developer of the model in tensor flow is used to detect which values best fit the model so there is a good set of hyper parameters to start the training.

Nodes are then created for the small model that is to be trained which will be the initial process in the detection process. Nodes are then added thereafter to the initial once the learning process is started. The session object is the interface to run the graph that helps in identifying from which node the processing has to run and from which the output will have to be retrieved i.e. which set of values from the graph are the best fit for the model being used.

This results in something called as tensor objects which in this case will only be a single object. This object might be thought as a 3 X 299 X 299 multi-dimensional matrix – three dimensional in this case since a standard image of 299 X 299 pixels with RGB values has been considered. These values will be used to identify a face in the image frame to further make a decision on a match with the existing trained faces. If there isn't any face detected, then there wouldn't be any notification thrown to the user. Else, there will be a notification pushed to the user. If there is a face detected in the image, it will output and send the result in the form of an XML or JSON with the image and the boundary parameters.

The bandwidth is also kept in mind while designing this approach. For example, each and every frame is not sent to the cloud for identification by the backend model. Rather, the initial check is done in the microcontroller module and if the object detection section finds something, the corresponding frames are sent to the cloud for further check. And now, when it detects an unidentified face, it sends a notification to the user notifying him of a potential breach.

VI. Results And Discussion

The proposed system was tested as a model of smart security. The proposed security system detects whether the user is at the secured area and accordingly activates the system. There are various parameters which can be adjusted in this software such as the kind of photos to capture and analyzed by the data analysis algorithms and mobility of the modules placed at the area which can be moved with high flexibility.

The developed security system has a good response time to the sensor and sends notifications to the application and a message on the phone when it detects any breach in terms of a fire when the sensor value is increased above a desired level or upon the detection of any form of intrusion which is identified by the photo sensor and the camera modules that interact with the microcontroller. The time taken by the system to deliver the message to the user is dependent on the coverage area or range of the specified mobile network. It also sends a notification to the user using a mobile application via the internet. The choice of the internet or GSM network is made depending on their availabilities keeping in mind the cost incurred to the user and the user is notified accordingly in about five to ten seconds while using the proposed system.

The computer vision algorithm designed produces a final image as shown in Fig. 4. Any new objects identified by the computer vision algorithm are highlighted using red and green rectangles as shown in Fig. 1. As an extension to this output, these images along with their parameter values, when sent to the server, are passed as input to the trained convolutional neural network present on a high capacity computer equipped with a graphics processing unit (GPU). The model simply outputs whether the image passed to it as input matches any other images contained in its database by checking the input against the preset list of users that is configured in the system during the initial setup when the neural network is trained with this list. The addition of any new users requires the system to be retrained with new information.

Advantages of Proposed Work:

1. The use of Image processing for object detection and Machine learning for photo analysis helps us achieve maximum security with least risk.

2. The system can check the status of the secured area and detect the faces in the area at the time of the breach and notify the user both via the cellular network as well as the internet.

3. The use of computer vision and Image processing helps identify objects frame by frame using improved version of the open source OpenCV's computer vision algorithms to achieve higher accuracy when compared to the trivial systems.

4. The system can be improved to use the open project FaceNet model to increase the achieved accuracy.

5. Deployment of this system is very simple since a very light set up is installed at the place to be secured leaving the rest of the computation to happen on the cloud.

6. The success rate for such a system is directly proportional to the accuracy of the artificially intelligent model used to verify the captured images. As a result, an improvement in the machine learning model will have a multifold increase in the effectiveness of the proposed system.

7. The use of smart phones as client side devices eliminates the use of additional equipment. Additionally, fundamental features like GPS and GSM are used to incorporate location tracking into the system.

Disadvantages:

1. In case of the false positives and false negatives not being handled properly, the system could trigger a false alarm.

2. The dependency on mobile phone devices at the user's end could result in an event where the user would not be notified of a breach in case of unreachability of the device through both GSM and internet.

3. The addition of new users to verify the input to the model against, the model would have to be retrained to incorporate these changes.

VII. Conclusion And Future Work

Improvement to this system can be done using the OpenFace and the classifier it offers. OpenFace helps us to get the 128 measurements of the face and that is sent as an input to the classifier. Looking at all the measurements of the images which are measured before and the classifier will check with the closest match of the face. This can be further enhanced using the FaceNet model of Google which can produce better results. FaceNet was able to produce an accuracy of about 99.63%. The loss function used to minimize the error is as follows-

$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+ \qquad (3)$$

The above function represents the embedding in a multi-dimensional space, where *x* represents the image in the function. The loss is calculated according to the nearest neighbor classifier. This loss function here tries to reduce the distance between the similar images x^a_i and x^p_i and away from the other images $\mathrm{x}^{\mathrm{n}}_{\mathrm{i}}$. Here α is the margin enforced between the positive and negative images.

The algorithm can be further improved using the blob detection algorithm which aims to detect the areas in images that differ in any property or similar in property. The system can be further enhanced using a new way of memorizing the faces of the people that newly visit the area to be secured which would result in the neural network model to be automatically retrained to adapt to the changes that result during the addition of new images. This also avoids the need for deployment on a server with extremely high computational power since the cost of training after the initial setup is much lesser than initial cost.

The vicinity-based security system has been designed and tested with the mobile network. The user can get alerts anywhere through the GSM technology or using the Wi-Fi shield thus making the system location independent. A flexible way to control and explore the services of the mobile and the network commands is used in the system. The automatic activation and deactivation of the system is supported using the user's location using the end user application which notifies system to activate itself. The trained convolutional neural network model that contains model parameters must be secured to the highest extent as a loss of this information could be very costly.

The system is able to check for the status of the secured area with the help of the camera setup bundled with the core hardware microcontroller module and when there is a breach detected by the security system, the images are to the cloud to check for the faces already in the system and also store the newly added images to the database to record any possible breach.

REFERENCES

[1] Lawrence, Steve, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. 1997. "Face Recognition: A Convolutional Neural Network Approach." IEEE Transactions on Neural Networks, Volume 8; Issue 1. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=554195C. Gomez and J.Paradlls, "Wireless home automation networks: A survey of architectures and technologies", IEEE Communications Magazine, Vol.48, No.6, pp.92-101, 2010

[2] J. Bangali and A. Shaligram, "Design and Implementation of Security Systems for Smart Home based on GSM technology", International Journal of Smart Home Vol.7, No.6, pp.201-208, 2013.

[3] Marcin Andrychowicz, Misha Denil, et al., "Learning to learn by gradient descent by gradient descent", arXiv preprint arXiv:1606.04474v2 [cs.NE] 30 Nov 2016.

[4] A. Antony and Prof. G. R. Gidveer, "Live Streaming Motion Detection Camera Security System with Email Notification using Raspberry Pi", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Special Issue - AETM, pp.142-147, 2016.

[5] M. W. Ren, J. Y. Yang, and H. Sun, "Tracing boundary contours in a binary image[j]," Image and Vision Computing, vol. 20, pp125-131, 2002.

[6] J. Rao, J. Lin, S. Xu, and S. J. Lin, "A new intelligent contour tracking algorithm in binary image," in Proc. 4th International Conference on Digital Home, 2012, pp 18-22.

[7] G. Pradeep, B. S. Chandra and M. Venkateswarao, "AdHoc Low Powered 802.15.1 Protocol Based Automation System for Residence using Mobile Devices", IJCST Vol.l. 2, No.1, pp.93-96, December 2011.

[8] Live streaming DIY system [Online] Available-http://www.networkworld.com/article/2925722/security0/home-security-demystified-how-to-build-a-smart-diy-system.html

[9] Angela Antony, Prof. G. R. Gidveer, "Live Streaming Motion Detection Camera Security System with Email Notification using Raspberry Pi" IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Special Issue - AETM'16, pp.142-147.

[10] Aldrich D'mello, Gaurav Deshmukh, Manali Murudkar and Garima Tripathi, "Home Automation using raspberry pi 2", International Journal of Current Engineering and technology, vol. 6, no.3, May, 2016

[11] Bhavik Pandya, Mihir Mehta, Nilesh Jain, Sandhya Kadam "Android Based Home Automation System using Voice Commands and Bluetooth" International Research Journal of Engineering and Technology Vol. 3, Issue 04, April, 2016.

[12] Pradeep.G, B.Santhi Chandra, M.Venkateswarao, "AdHoc Low Powered 802.15.1 Protocol Based Automation System for Residence using Mobile Devices", Dept.of ECE, K L University, Vijayawada, Andhra Pradesh, India IJCST Vo l. 2, SP 1, December 2011.

[13] P.J. Philips, H. Moon, S.A. Rizvi, and P, J. Rauss, "The FERET Evaluation Methadology for Face-Recognition Algorithms," IEEE Trans. *Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.

[14] A. Nikolaidis and I. Pitas, "Facial Feature Extraction and Determination of Pose," *Pattern Recognition,* vol. 33, pp. 1783-1791, 2000.