International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

DEVELOPING A FACE RECOGNITION SYSTEM USING CONVOLUTIONAL NEURAL NETWORK AND RASPBERRY PI INCLUDING FACIAL EXPRESSION

C.R Vimal Chand[1], Dr.G.P Ramesh Kumar[2]

[1] *Research Scholar and Associate professor in Computer Science Department, Sri Ramakrishna College of Arts And Science, Coimbatore,Tamilnadu,(India)*

(Formely S.N.R Sons College-Autonomous)

[2] *Associate professor in Computer Science Department, Government Arts College, Kulithalai, Tamilnadu, (India)*

ABSTRACT

Facial expression recognition systems have attracted much research interest within the field of artificial intelligence. Many established facial expression recognition (FER) systems apply standard machine learning to extracted image features, and these methods generalize poorly to previously unseen data. This project builds upon re-cent research to classify images of human faces into discrete emotion categories using convolutional neural net-works (CNNs). We experimented with different architectures and methods such as fractional max-pooling and fine-tuning, ultimately achieving an accuracy of 0.48 in a seven-class classification task. Detection of face is a system used to determine any faces in a captured image and, if present, gives the image location and content of each face. Security and surveillance are the two important aspects of human being. In this work we tried to construct face detection and recognition system that will be capable of processing images very fast when capturing with very high face detection rate with positive ratio. This paper describes a simple and easy hardware implementation of face detection system using Raspberry Pi. The system is programmed using Python programming language. A real time face detection and face recognition is carried out from specific images, i.e. Object Recognition, is find out and the proposed system is tested across various standard face databases ,Emotional face databases with and without noise and blurring effects.

Key words* : *CN N, Max-pooling, Base line classifier, Haar pattern, Python, Open CV, Raspberry Pi, LBP

I.INTRODUCTION

One of the unique features of our brain is that it can think only in images not in words. Once you may forget to keep your Car's key but you will never forget to bring a face with you. God has given everyone a unique face. Face is the most important part of our body, so that it can reflect many emotions of a person. From a long year ago, we are using non living thing (plastic cards, keys, PINS, smart cards, tokens) for authentication and to get grant access in restricted areas like NASA, DRDO and ISRO etc. There are two types of biometric as

physiological characteristics (face, fingerprint, finger geometry, hand geometry, palm, iris, ear and voice) and behavioral characteristics (gait, signature and keystroke dynamics). Sometimes your behavioural traits may changes because of illness, fear, hunger etc. Face detection and recognition system is more expensive, exact, easy to understand and non intrusive process as compare to other biometrics. The system will fall into two categories as face detection (1:1) and face recognition (1:N).In the face detection we have to classify between face versus non face region while in recognition process we have to compare that single face image with multiple images from the input image. While capturing an images from a web cam we have to come across some problems like pose (position of camera), presence of structural components (spectacles and beard), facial expression, occlusion (obstructed by someone), image orientation (variation in rotation), imaging condition (lightning and camera characteristics) etc. Face detection is a common feature of digital cameras since 2006. Automatic face detection and recognition system is placed at New Zealand airport only for that citizen since 2010 while airport in Europe started to be equipped with similar systems from 2008 for security purpose. The objective of this project is to classify images of hu-man faces into discrete emotion categories. Many estab-lished facial expression recognition (FER) systems use stan-dard machine learning and extracted features, which do not have significant performance when applied to previously unseen data [1]. Within the past few months a few papers have been published that use deep learning for FER [2] [13] which have been successful at achieving about .60 accuracy on the EmotiW and other publicly available data sets. Not-ing the success of CNNs in this domain, our objective is to experiment with both new and existing network architec-tures to achieve similar results on a new data set.Artificial intelligence systems to recognize human emo-tion have attracted much research interest, and potential applications of such systems abound, spanning domains such as customer-attentive marketing, health monitoring, and emotionally intelligent robotic interfaces.

II. LITERATURE SURVEY

In this paper internet of things has initiated tremendous growth in internet and products that are connected to internet. Internet of things components are cost effective, small in size computational power for application oriented components can be used in surveillance system by using open CV and python. Stored faces in cloud can be recognized [1].

The main strategy of this paper is taking attendance in organization etc. for these time periods are set after completion of time period. Attendance are directly stored this raspberry pi 2 module used. For high speed operation eigen face algorithm used but when a number of students faces increases the accuracy will decreases [2]. In this paper solve the problem of high image processing speed. The two viewpoints are face descriptor tool for face recognition and feature extraction by using computer vision, raspberry PI used to create the data base. Database can be used to match the input and then shown output of the screen [3].

A portable real time facial recognition system that is able to play personalized music based on the identified person's preferences was developed. The system called portable facial recognition JUKEBOX using fisher faces. Raspberry pi was used as the hardware. When implemented on a raspberry pi, the system is able to facial recognition. This paper presents a cost effective solution that is cross platform Linux OS and windows. Future

work of this project would be to future increases the precision of the system by improving algorithm to be more robust to other environments [4].

Social platform and social media age and gender classification has become applicable. to an application. In this paper use in convolutional neural systems and the end use in direct convolutional neural system which used in measure of learning data is limited [5].This paper propose that face recognition based on fusing the near infrared and visible images of face images with distributed sensing obtaining common component. Used distributed compressive sensing and innovation component of near infrared and visible image [6]. Unsoojang and Euichul lee proposed a methodology of pixel based face recognition method face recognition method are used in appearance based, texture based, geometry based. For this used in principle component analysis, linear discriminant analysis and local binary pattern [7].

Li Yong –Qiang and Panjin explain a one sample image recognition algorithm based on improved sub-pattern principle component analysis. For this face divide some blocks and sub pattern then PCA is used and recognize the face supports vector machine used. The results demonstrate that the proposed algorithm has improved the recognition rate and speed of face and has robust [8]. Maryam Moghaddam and Aped Meshiqini are presented automatic facial recognition for used the local directional pattern to represent the geometry and analyse the performances. Template matching and support vector machine are used and improved algorithm used entropy LDP+SVM [9]. Navin Prakash and Yashpal Singh presented support vector machine is the recent technology. This technology gives the decent broad view performance and SVM can be extended in many other ways and some ways SVM cannot gives better result than technology are extended in fuzzy support vector machine [10].

In this paper proposes an efficient head pose determination method. In face recognition variation related problem solve then detect face by using ad boost. Then pre-processing on detected face. For this PCA based face recognition approach depending on a single –pose face DB [11]. In this paper proposed that image capturing technique in an embedded system based on raspberry pi.an implement purpose embedded platform is very unique. In this use in image capturing and recognition algorithm. Raspberry pi and its peripherals and then actualized embedded image capturing using raspberry pi system [12].

In this paper G. Ghinea propped the novel methodology to recognize the face images. For this use in Yale face database and ORL face database and gradient orientation approach are used and also used in schurvectors for subspace learning in future to improve the accuracy of face recognition [13]. Mounika B. R, Reddy proposed a Neural Network based face detection by using Gabor Features. The complex classifier are used to better the algorithm by representing Gabor faces use in distance measures in future improve the detection rate and speed of the detection process [14].

Tudor BARBU proposed a novel approach of face recognition which based on Gabor filtering and supervised classification. The 2D filter bank are used and then produces 3D robust face for vector average distance used in supervised classifier and threshold based face verification method used by using this technique a high facial recognition rate is obtained [15].

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

III. METHODOLOGY

3.1 Webcam

A webcam is video cameras that feeds or stream its image in real time to or through a computer to a computer network. a webcam is generally connected by a USB cables, or similar cables built into computer hardware such as laptop. Popular uses include security surveillance, computer vision and recording video.

3.2 Haar Like Feature for Face Detection

Haar like features are digital image feature used for object detection but here we used it for face detection. The biggest advantage of it over most other features is its calculation speed. Fig. 1 shows the types of Haar like feature. Generally eye region is darker than other region from the face. Figure 3 shows how Haar like feature is used for face detection purpose. the complete preprocessing steps, which includes binary to gray scale image conversion, Histogram Equalization method (HE), Laplacian of Gaussian filter (LG) and final step is contrast adjustment. Preprocessing is done because we have to remove influence cause by illumination variation for accurate face recognition.

• Edge feature

• Line feature

• Center-surround feature

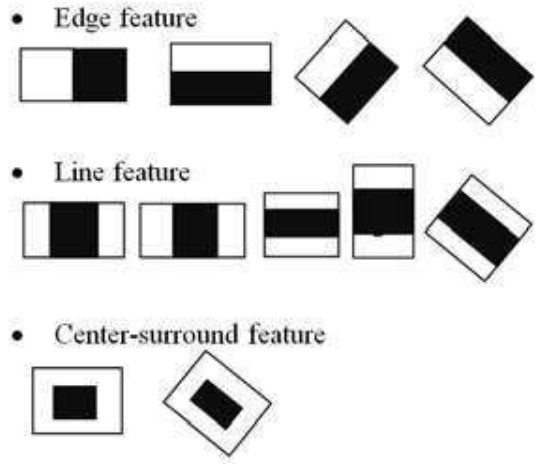


Figure 1: Types of Haar Like Features

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

3.3 WORK RELATED WITH FACE EXPRESSION AND EMOTION

Yu and Zhang achieved state-of-the-art results in EmotiW in 2015 using CNNs to perform FER. They used an ensemble of CNNs with five convolutional layers each [14]. Among the insights from their paper was that randomly perturbing the input images yielded a 2-3% boost in accuracy. Specifically, Yu and Zhang applied transforma-tions to the input images at train time. At test time, their model generated predictions for multiple perturbations of each test example and voted on the class label to produce a final answer. Also interesting is that they used stochastic pooling rather than max pooling because of its good perfor-mance on limited training data.

Kim et al. achieved a test accuracy of .61 in EmotiW2015 by using an ensemble based method with varying network architectures and parameters [2]. They used a hierarchical decision tree and an exponential rule to combine decisions of different networks rather than simply using a simply weighted average, and this improved their results. They initialized weights by training networks on other FER data sets and using these weights for fine-tuning.

Mollahosseini et al. have also achieved state of the art results in FER [1]. Their network consisted of two convolu-tional layers, max-pooling, and 4 Inception layers as intro-duced by GoogLeNet. The proposed architecture was tested on many publically available data sets. It received a lower test accuracy of 0.47 on the Emotion data base

Figure 2. A sample of images from the data set, labeled with their corresponding emotions.

the-art test accuracies on other data sets (i.e. 0.93 on CK+). When compared to an AlexNet architecture, their proposed architecture improved results by 1-3 percent on most data sets.

In [5], Graham proposes a specific type of stochastic pooling, called fractional max-pooling, that achieves the regularization effect of standard max-pooling without dis-carding as much spatial information at each pooling step. As this method is most suited for data sets with small im-ages, we experimented with it on our data set.

Instead of reimplementing published networks, we de-cided to take the key insights from these papers and experiment with different networks. All of the proposed papers used a network about 5 - 7 layers deep and image perturbation.

3.4 Data

We trained and tested our models on the data set from the Kaggle Facial Expression Recognition Challenge, which comprises 48-by-48-pixel grayscale images of human faces, each labeled with one of 7 emotion categories: anger, dis-gust, fear, happiness, sadness, surprise, and neutral. We used a training set of 28,709 examples, a validation set of 3,589 examples, and a test set of 3,589 examples.

As illustrated in Figure 1, the data set's images vary con-siderably in scale, viewpoint, and illumination.

We note that that all of the images are preprocessed so they form a bounding box around the face region. However, there are differences in angle, lighting, and objects – for example, some faces are adorned with glasses or covered by long hair.

IV.TECHNICAL WORK

We implemented three different classifiers from scratch:

(1) a baseline classifier with one convolutional layer, (2) a CNN with a fixed size of five convolutional layers, and a (3) deeper convolutional layer with parameterized depth and filtered features. Filters are used to remove noise and extract features. The above CNN architectures are experimented and compared with AlexNet architecture.

A deeper CNN with a paramaterizable number of con-olutional layers, filter dimensions, and number of filters. For each of these models, we tuned parameters including learning rate, regularization, and dropout. We also exper-imented with using batch normalization [6] and fractional max-pooling [5].

Finally, we implemented multiple classifiers using fine-tuning with variations on the number of layers retained, the number of layers backpropagated through, and the initial network used. We experimented with fine-tuning using two existing models: (1) VGG16, the model from Caffe's Model Zoo which was trained on ImageNet, and (2) VGGFace, a network trained on a facial recognition data set.

4.1. Baseline classifier

We implemented a baseline softmax classifier using fea-tures from a single convolutional layer. The architecture was one convolutional layer, followed by one fully con-nected layer, and a final softmax layer. The initial baseline did not use any regularization, dropout, or batch normaliza-tion.

4.2. Five-layer CNN

We implemented a first-pass CNN with a fixed depth of five convolutional layers. The model was trained using the architecture and was trained using the following characteristics.

Parametrized dropout rate, learning rate, and l2 regu-larization
Batch normalization (optional) after each layer Adam update rule
Weight initialization for using ReLU nonlinearities as presented by He et al.
3x3 convolutional filters with stride 1 and zero-padding to preserve spatial size
2x2 max pools with a stride of 2
Using this model, we also experimented with both traditional max-pooling and fractional max-pooling (described below).

4.3. Deeper CNN

The architecture of our deeper CNN is outlined in Ta-ble 2. We use the same network characteristics as the five-layer CNN with additional network structure parame-ters. With parameterizable layer depths and filter sizes, this model has a greater possible capacity than the five-layer

4.4. Fractional max pooling

Traditional max-pooling is often applied using 2x2 re-gions with stride 2. This configuration discards 75% of the data, and such drastic reduction in spatial size can be unde-sirable, particularly in data sets with small images such as ours.
Fractional max-pooling (FMP) is "gentler" in that it al-lows for a reduction of the spatial size of a layer by any arbitrary fraction. When the fraction is between 1 and 2, pooling regions of size 1x1, 1x2, 2x1, and 2x2 are stochas-tically shuffled to achieve the desired output size.
More specifically, FMP is divides a matrix into either disjoint or overlapping pooling regions as follows, and as described in more detail in [5].

$P = [a_{i\,1}; a_i\ 1] \times [b_{j\,1}; b_j\ 1]$ or $P_{i;j} = [a_{i\,1}; a_i] \times [b_{j\,1}; b_j]$.

We obtain the these sequences by taking a random permutation of an appropriate number of ones and twos which de-note the types of pooling regions, and then pooling within these stochastically determined regions.
We expected FMP to be useful since our data sets im-ages are small. Using FMP would allow us to form deeper networks without as rapid data loss, at the expense of more computation. We experimented with random, overlapping fractional max-pooling as described in [5].

4.5. Fine tuning

We fine-tuned using the model VGG16 as listed in Caffe's Model Zoo and described in [7]. VGG16 was trained on ImageNET [11], which has data that is quite different than ours in content. VGG16 was trained to per-form object detection and localization on images of various objects (not faces). Since the objective was different from ours but the subset of the ImageNET data set it was trained on was in the same order of magnitude as the size of our data set, we tried the following two configurations of fine-tuning: (1) using the lower few layers of the

VGG16-trained network, and (2) using the entire VGG16-trained network. With both of these configurations, we trained two fully con-nected layers at the end (on top) of the network.

In addition, we found a network trained on face images

– VGGFace – available from the Visual Geometry Group at the University of Oxford [12]. This network was trained on a very large-scale data set (2.6M images, 2.6k people) for the task of face recognition. Since the data set was trained for a similar application but on a much larger data set than ours, we tried fine-tuning with the following configuration:

4.6 Training for face Recognition Using Raspberry PI

USBCam connect to Raspberry PI. For image acquisition camera through continuously read the frame. Means acquire the image. For preprocessing image are going to standard size. The gray scale converter are used. They check only two values black and white, lesser time are required. Then check the detect faces used Haar classifier. Whether faces or object are not. Region of interest detect the faces.

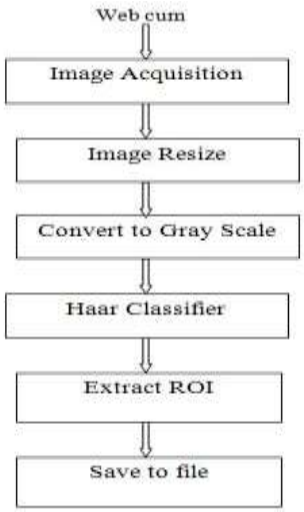


Figure 3

4.6.1 Face Recognition

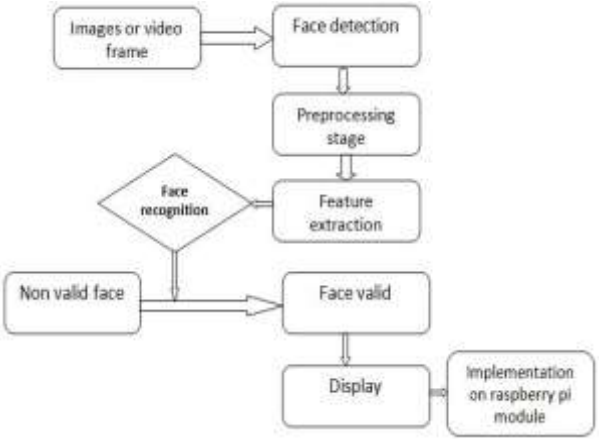
The detected faces check the data base and then compare LBPH value and then match the present data base and face are recognized.

4.6.2 Local Binary Patterns

Local binary patterns (LBP) are a type of visual descriptor to automatically classify and identify textures and patterns in images. Local binary patterns are a type of visual descriptor used for classification in computer vision.

Figure 4: Block Diagram for Face Detection and Recognition

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354



4.3 Advantages

1 It is used in Real time processing.

2 It is cost effective.

3 LBPH is efficient as compared to PCA.

4.4 Limitation

If number of data base are increase then frames iteration are increase and hence system will work slow.

4.5 Application

1. Employment Attendance system.
2. Student Attendance system.
3. Social Networking.

The aim is to make a smart surveillance system using Raspberry Pi along with PIR sensor and Raspberry Pi-Camera. PIR sensor is used to detect the motion whenever someone comes within its range. As soon as PIR Sensor detects the motion, Pi-Camera activates and captures an image. This image is then stored in the system and finds for a human face in the captured image using OpenCV and Python. The detected human face is then compared with the faces stored in

the database using Local Binary Pattern algorithm. If the face matches with the ones in the database no alert is generated else host gets a text message on his android mobile phone via Pushetta application using Wifi adapter. In this way, this system helps to identify only unauthorised persons. This helps to overcome the drawback of CCTV and Motion Detection systems which only monitor or alert host based on he motion detected whether it is authorised person or not.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

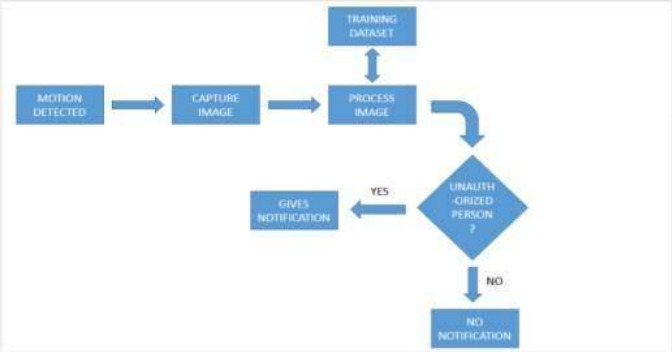
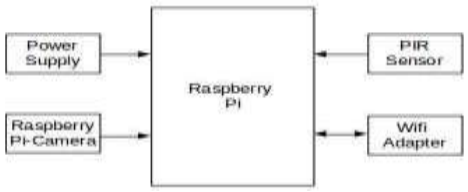


Figure 5: Block Diagram of Processing System.

V.HARDWARE IMPLEMENTATION

The system consists of Raspberry Pi, PIR Sensor, Raspberry Pi-Camera, Wifi adapter and Power Supply.

Figure.6 Block diagram for
Hardware



A. Raspberry Pi

Raspberry Pi is a low cost (35 dollar), credit card sized, computer that performs various applications. Some of its main features include 1GB of RAM, 4 USB Ports, General Purpose Input Output pins, Linux support. These features gives programmers a wide range for diverse applications.

B. PIR Sensor

Passive Infrared Sensor is an electronic sensor that measures infrared lights radiating from the objects in its field of view. Its maximum range is about 10m which is suitable for motion detection applications.

C. RaspberryPi-Camera

Raspberry Pi-Camera module is a 8MP camera with full HD recording capability. This provides a perfect solution for face recognition.

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

5.1 Software Implementation.

This system is largely based on Python programming from detecting the motion to generating an alert. Various Python libraries are used to control PIR Sensor for detecting the motion, Python is used for Pi-Camera to capture and process

images. The captured image is then processed using OpenCV library that integrates with Python. The Face-recognition part is carried out by the Local Binary Pattern(LBP) algorithm.

A. OpenCV

OpenCV (Open Computer Vision) is a library mainly aimed at real-time computer vision. It provides great support for face detection and face-recognition techniques using Python.

B. Local Binary Patterns

Local binary patterns (LBP) is a type of algorithm used for classification in computer vision. It has been found to be a powerful feature extraction and classification purposes.

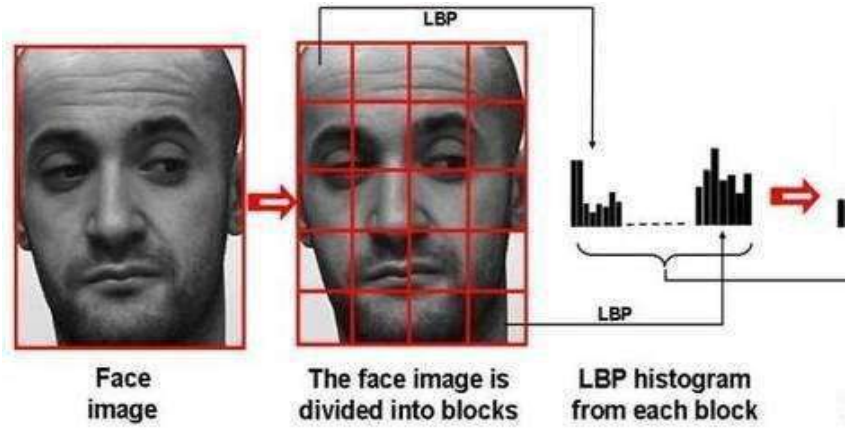


Figure 7 Intensity calcuations in Local Binary Pattern

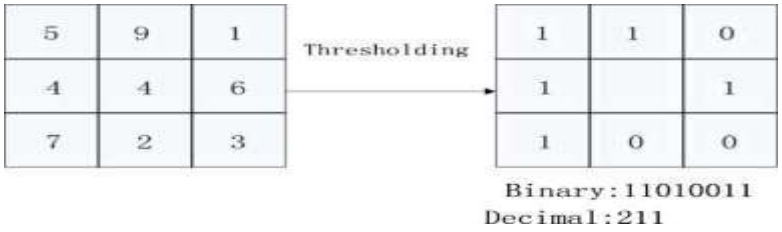


Figure. 8 Local Binary Pattern algorithm working

VI.CONCLUSION

In this project, we addressed the task of facial expression recognition and aimed to classify images of faces into any of seven discrete emotion categories that represent univer-sal human emotions. We experimented with various tech-niques, such as fine-tuning and fractional max-pooling, and achieved our highest accuracy (0.48) on a CNN trained from scratch with seven convolutional layers. Given more time, we would have liked to combat overfitting and approach state-of-the-art accuracies of around 0.61. Surveillance system provides an efficient way for monitoring suspicious activities. Traditional systems are efficient and have low maintenance cost. However, energy consumption is more as the system is continuously powered on. Proposed IOT based smart surveillance system provides energy management by turning the system ON, based on the occurrence of a particular motion. System will sense the motion and depending on the detected motion system will switch on the camera, capture the image of intruder, recognize it and send a notification on owner's smart phone if the person is not recognized by the system.

REFRENCES

[1] Ms. Naga Jyoti and Mr. K. Vijaya Vardhan, “Design And Implementation Of Real Time Security Surveillance System Using IoT”, Communication and Electronics Systems (ICCES), International Conference: IEEE, 2016.

[2] M. Surya Deekshith Gupta, Vamsikrishna Patchava, and Virginia Menezes: “Surveillance and Monitoring System Using Raspberry Pi and SimpleCV”: Green Computing and Internet of Things (ICGCIoT), IEEE, 2016.

[3] Aruni Singh, Sanjay Kumar Singh, Shrikant Tiwari, “Comparison of Face Recognition Algorithms on Dummy Faces”, The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.4, August 2012.

[4] Aamir Nizam Ansari, Mohamed Sedky, Neelam Sharma, Anurag Tyagi, “An Internet of Things Approach for Motion Detection using Raspberry Pi”, International Conference on Intelligent Computing and Internet of Things (IC1T), 2015.

[5] R.Chandana, Dr.S.A.K.Jilani, Mr.S.Javeed Hussain, “Smart Surveillance System using Thing Speak and Raspberry Pi”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 7, July 2015.

[6] Sanjana Prasad, P.Mahalakshmi, A.John Clement Sunder, R.Swathi, “Smart Surveillance Monitoring System Using Raspberry PI and PIR Sensor”, (IJCSIT) International Journal of Computer Science and Information Technologies, 2014.

[7] Huu-Quoc Nguyen, Ton Thi Kim Loan, Bui Dinh Mao, Eui-Nam Huh, “Low cost real-time system monitoring using Raspberry Pi”, IEEE, 2015

[8] Sarabjit Singh, Amritpal Kaur, T aqdir, “A Face Recognition Technique using Local Binary Pattern Method”, International Journal of

Advanced Research in Computer and Communication Engineering, 2015.

[9] Sushma Jaiswal, Dr. (Smt.) Sarita Singh Bhadauria , Dr. Rakesh Singh Jadon, " Comparison between face recognition algorithm-eigenfaces, fisherfaces and elastic bunch graph matching ", Journal of Global Research in Computer Science, 2011

[10] https://en.wikipedia.org/wiki/Image_processing

[11] https://en.wikipedia.org/wiki/Local_binary_patterns

[12] [12] Ekman and W. V. Friesen. Emotional facial action coding system. Unpublished manuscript, University of
California at San Francisco, 1983.

[13] B. Graham. Fractional max-pooling. 2015.S.
Ioffe and C. Szegedy. Batch normalization:
Accelerating deep network training by reducing
internal covariate shift.JMLR Proceedings,
2015.

[14] M. S.-S. S. M. B. Z. L. X. S. B. P. J. Zhang, S.
Ma and Mech. Salient object subitizing. IEEE
Conference on Computer Vision and Pattern
Recognition (CVPR), 2015.

[15] R.Y. Jia, E. Shelhamer, J. Donahue, S. Karayev,
J. Long, R. Gir-shick, S. Guadarrama, and T.
Darrell. Caffe:Convolu-tional architecture for
fast feature embedding. arXiv preprint
arXiv:1408.5093, 2014.

[16] Lab41. Misc, caffekeras util. https://github.com/Lab41/Misc/blob/master/blog/kerasvgg/ caffekeras_util.py, 2015.Lasange. Lasange, fractional pooling layer pull re-quest. https://github.com/Lasagne/Lasagne/pull/171, 2015.