Information Sciences 479 (2019) 416–431

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services

Muhammad Sajjad[a], Mansoor Nasir[a], Fath U Min Ullah[a], Khan Muhammad[b], Arun Kumar Sangaiah[c], Sung Wook Baik[b,*]

[a] *Digital Image Processing Laboratory, Department of Computer Science, Islamia College Peshawar, Pakistan*
[b] *Intelligent Media Laboratory, Digital Contents Research Institute, Sejong University, Seoul, Republic of Korea*
[c] *School of Computing Science and Engineering, Vellore Institute of Technology, Vellore-632014, India*

ARTICLE INFO

Article history:
Received 21 August 2017
Revised 23 June 2018
Accepted 8 July 2018
Available online 17 July 2018

Keywords:
Facial expression analysis
Machine learning
Smart security
Suspicious activity recognition
Smart cities

ABSTRACT

Facial expression recognition is an active research area for which the research community has presented a number of approaches due to its diverse applicability in different real-world situations such as real-time suspicious activity recognition for smart security, monitoring, marketing, and group sentiment analysis. However, developing a robust application with high accuracy is still a challenging task mainly due to the inherent problems related to human emotions, lack of sufficient data, and computational complexity. In this paper, we propose a novel, cost-effective, and energy-efficient framework designed for suspicious activity recognition based on facial expression analysis for smart security in law-enforcement services. The Raspberry Pi camera captures the video stream and detects faces using the Viola Jones algorithm. The face region is pre-processed using Gabor filter and median filter prior to feature extraction. Oriented FAST and Rotated BRIEF (ORB) features are then extracted and the support vector machine (SVM) classifier is trained, which predicts the known emotions (Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise). Based on the collective emotions of the faces, we predict the sentiment behind the scene. Using this approach, we predict if a certain situation is hostile and can prevent it prior to its occurrence. The system is tested on three publically available datasets: Cohen Kande (CK+), MMI, and JAFEE. A detailed comparative analysis based on SURF, SIFT, and ORB is also presented. Experimental results verify the efficiency and effectiveness of the proposed system in accurate recognition of suspicious activity compared to state-of-the-art methods and validate its superiority for enhancing security in law enforcement services.

1. Introduction

A major communicative source and common exchange of feelings in daily communication is facial expression. Facial expressions and other gestures are able to deliver these cues as non-verbal face-to-face communication. These cues can deliver a complete understanding of the desired message. Facial expression has high weightage over the words that are being spoken during our personal exchange of ideas. Recently, researchers have suggested developing such robust and dedicated devices that helps recognise the emotions and different moods [37]. Different techniques have been applied to develop

[*] Corresponding author.
E-mail addresses: Muhammad.sajjad@icp.edu.pk (M. Sajjad), khanmuhammad@sju.ac.kr (K. Muhammad), sbaik@sejong.ac.kr (S.W. Baik).

https://doi.org/10.1016/j.ins.2018.07.027
0020-0255/© 2018 Elsevier Inc. All rights reserved.

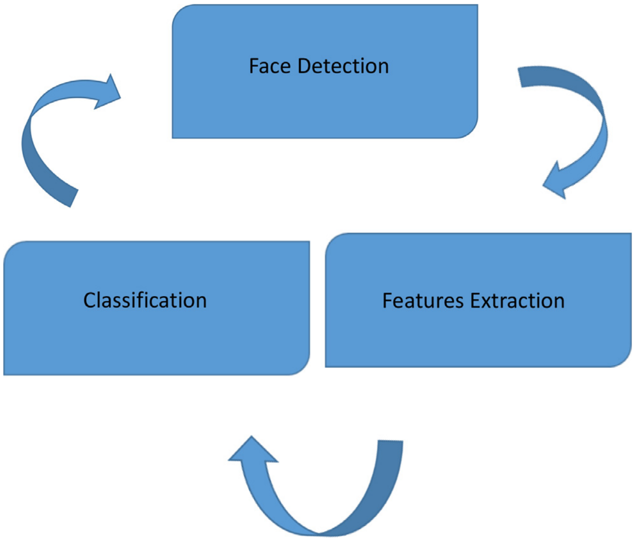


Fig. 1. Major Three steps of FER.

automated tools for facial expression recognition (FER) that are applied during interviews [14], surveillance systems [22], aggression detection [3], and expression recognition while driving [45]. The research in facial expression recognition plays an influential and vital role in the field of machine learning where intelligent robots and smart cameras are trained to analyse the mood of humans. Emotion recognition and related technologies have many applications in our day-to-day life such as lie detection. Therefore, the research in facial expression recognition is becoming a multi-disciplinary research area. Along with other applications, FER has vast applications in the field of security, i.e., identification and verification of a person's expressions in image/video frames. Identifying humans from faces is what we humans do implicitly, and we are quite good at it. But when it comes to computers, it is not an easy task for image processing applications, as there are several parameters that need to be calculated precisely before recognising emotions.

Over the past several years, the emotion recognition technology grabbed the attention of an overwhelming number of researchers and it is slowly replacing other biometrics. This is mainly because it enables us to record at a distance without interacting with the subject itself, making it convenient for a wide range of applications. Emotion recognition systems range from criminal detection in a national database to social media websites; and they also can be used for the identification of suspects on the international borders [34]. This issue is still attracting many researchers. More recently, Goyal et al. [18] and Allaert et al. [2] presented a comprehensive survey related to the current advancements in the field of Facial Expression Recognition. Smart cities are not a far-fetched prospect anymore. Tech giants are investing billions of dollars to make it a reality. Along with other facilities, smart cities needs to be able to provide the best possible security for their citizens. Providing a secure and defensive environment for ensuring security under smart cities is a challenging task. The proposed system is designed to ensure security enforcement in smart cities using facial cues of suspects. Faces convey messages implicitly and it cannot be controlled in a tense situation. Hence, it is ideal for security agencies to identify suspects using facial cues and recognising expressions in real-time. Usually, a trained psychiatrist is observing the suspect from a distance, and not all cues can be observed efficiently – either by machines or by humans [42]. In most cases, a hostile situation is instituted when some unwanted situation is observed –j; like, for instance, a robbery or a quarrel between persons. In this case, a group comprising of some individuals may be in an anger or fear condition which is the sign of a suspicious event that is to occur during this scenario. In the proposed system, the features of each expression are calculated using Oriented Fast Rotated BRIEF (ORB) descriptor to train the Support Vector Machine (SVM) classifier. Instead of the entire image, only the calculated features are transmitted to the cloud and the emotion of each individual is recognised in a frame and the overall emotion of the entire image is calculated. Based on the overall expressions of the frame, the hostile situation is predicted and a pre-defined message is broadcasted against the activity.

Almost all FER-based systems are comprised of three steps. In the first step, the face is detected in a video stream, and is cropped as region-of-interest for the next step. After cropping the face region, it is resized into specified dimensions so that all the images given to the model become the same size. In the next step, some low level or high level or both features are extracted from the cropped region. In the last step, the features are classified using a classifier. In our case, the SVM classifier is used to train and predict the detected emotion. This system usually requires huge numbers of sample images to train the model. An overview of the main steps of the FER system is given in Fig. 1.

Emotion recognition is closely related to facial expression recognition, and still there is a research gap on FER-based systems. However, remarkable work is still in progress to advance it to the higher stage and to achieve every desired level of accuracy. The FER-based system itself can be appraised as a special case in the area of pattern recognition and in all those areas and techniques relevant to this. While designing the FER systems, we utilise the benefits of all these resources and various algorithms as building blocks in the same system. So the major components from all these resources and works are to determine an optimal combination of these algorithms. For this purpose, we divided the whole system into four modules, i.e., image preprocessing, face alignment, face feature extraction, and classification. To this end, different methods were implemented for each of the modules; and by comparing the performance of different combinations of feature descriptors, an optimal configuration was computed and selected which suits the desired resource constraint device such as Raspberry Pi. Our major contributions are listed as follows:

- A novel Raspberry Pi-based cloud assisted framework for facial expression recognition to assist suspicious activity recognition using a cost-effective solution for smart city security is presented.
- The Viola Jones-based face detector is incorporated in the proposed framework, providing efficient face detection results compared to state-of-the-art schemes. We proved that the accuracy of the system is much improved after the face is properly aligned. For this reason, we used a boost-regressor face alignment technique before extracting features. This technique promises efficiency with very little overhead.
- The proposed framework used ORB features to train an SVM model which is trained on cloud to recognise emotions. Instead of sending the entire face image, only extracted features are transmitted to the cloud, saving the transmission energy and bandwidth. We calculated the hash of each vector using the SHA-128 algorithm. This created a 128-bit hash for a given vector and then it is transferred to the cloud implicitly, along with the feature vector to provide integrity and confidentiality.

The rest of the paper is organised as follows: Section 2 reviews the related works and the proposed work is explained in Section 3. Experimental results and discussion are given in Section 4. The conclusion and future research directions are given in Section 5.

2. Literature review

In this section, we discuss some state-of-the-art techniques presented in the literature for FER. The first automatic facial expression recognition system was introduced in 1978 by Suwa et al. [38]; this was able to analyse facial expressions by tracking the motion of 20 identified spots in an image sequence. Facial expressions or facial cues can be used in many situations. For instance, Sajjad et al. [33] presented a framework where facial expressions can be used to retrieve relevant contents of a video stream where features were extracted using the histogram oriented gradient (HOG) descriptor with uniform local ternary pattern (U-LTP). Facial expression recognition is not new and various researchers proposed Facial Action coding systems. For example, Ekman et al. [15] proposed a system of facial active coding, which has the capability of perceiving various emotions via the facial muscles contraction and their relaxation both individually and simultaneously. They showed that there are different muscle motions which are treated as action units that can be used to track expressions where every unit is mainly the composition of some digits and letters. The combination of these action units can present various emotions to define the moods of the individuals. Yang et al. [46] proposed a technique which makes use of PCA [23]. Kshirsagar, V. P., M. R. Baviskar, and M. E. Gaikwad used the Eigen faces technique using linear discernment analysis [47] and Hidden Morrow Model [12] for emotions recognition and features extraction. More recently, the use of ANN [35] and CNN [27] provides fairly better results in both emotions detection and recognition.

The first and major step of all FER systems is the face detection and it is still a challenging task due to a lot of problems like image compression artifacts, high illumination, or low resolutions etc. For this purpose, researchers are constantly proposing to develop robust approaches. For instance, Zhu et al. [50] used the Contextual Multi-Scale Region-based Convolution Neural Network (CMS-RCNN) approach which accurately detects the face even if inverted or located at a very bleak angle. Lee et al. [24] has attempted to detect faces via skin color; for this purpose, he used the YCbCr model. The subcomponents of the underlying face, such as the mouth and eyes detected were considered as regions of interest (ROI). BRIEF points were extracted from those ROIs by applying a Bezier curve. The end result was an accurate face detection scheme, but the down side was that it is an expensive solution in terms of memory consumption and is not suitable for smaller devices such as Raspberry Pi. Al-Shabi et al. [1] integrated SIFT and CNN features for facial expression recognition. The authors claim that the proposed system can be trained with fewer numbers of images to achieve higher accuracy. A face recognition system based on the scale invariant feature transform (SIFT) [48] algorithm is used for the feature extraction of faces. These features are robust and fast but are weaker and have more time complexity than SURF [21].

Pantic and Rothkrantz [30] and Fasel and Luettin [16] reviewed the emotions recognition techniques for front-view images closely, along with the summarisation of mainly three components of emotions recognition which are face detection, emotions feature extraction and classification. For extracting image features, recently [49] a method has been proposed where the 2D image features are extracted by 1-norm regularised 2D neighborhood preserving projection (2DNNP) technique, where the neighborhood preserved features are extracted by 2DNNP, while Datta et al. [13] concatenate the geometric and texture-based features for facial emotion classification. They also used these features separately, but the overall performance of the concatenated features was more accurate during their comparison; however, facial feature localisation still remains a challenge.

The Facial expression classification is still an active research area and a lot of research is still on the peak, using different algorithms like Uçar et al. [40] used an algorithm where the classification of the facial expressions is done using

Table 1
Description of symbols and parameters of the proposed framework.

đ→ Dataset	L_c → Labels of the classifier
đ$_{\mathrm{TR}}$ → Training dataset	I_{CR}^{TR} → Cropped region of the training image
đ$_{\mathrm{CR}}$ → Dataset for cross validation	$\mathrm{I}_{(3N/5)+10}^{\mathrm{TR}}$ → 70% of the training images
$\mathrm{I}_1^{\mathrm{TR}}$→ Training image	C→ Classifier (SVM in our case)
I → represents the image	F_i^{ORB}→ ORB features of an image

curvelet transform which is applied to the region of the face through which the statistical features are extracted; the online sequential extreme learning machine (OSELM-SC) classifier is applied to obtained the features.

3. Proposed methodology

In this section, we present the details of the proposed framework and the explanation of each step is elaborated in the coming sections. Before starting, the dataset đ is divided into two parts, training set and testing set. The training set đ$_{\mathrm{TR}}$, which contains **N** number of images I, i.e. đ$_{\mathrm{TR}}$={$\mathrm{I}_1^{\mathrm{TR}}$, $\mathrm{I}_2^{\mathrm{TR}}$, $\mathrm{I}_3^{\mathrm{TR}}$**...** $\mathrm{I}_{(3N/5)+10}^{\mathrm{TR}}$} and the test dataset, i.e. cross validation test set đ$_{\mathrm{CR}}$ **=**{$\mathrm{I}_{(3N/5)+10}^{\mathrm{C}}$......**N**}. First of all, an image $\mathrm{I}_1^{\mathrm{TR}}$ is taken from the training set and its quality is enhanced using histogram equalisation and a median filter. For further removal of noise and blurring effects, we use Gabor Filter, which is discussed later in this section. An accurate and well-known algorithm of Viola Jones is used for face detection in the underlying image. We used the Viola Jones algorithm because of its simplicity, robustness and accuracy in terms of its use on Raspberry Pi. The latest face detection technique could be used to improve it further, but their performance on resource constraint devices are not yet assessed. For this demonstration, Viola Jones algorithms perform well with very little overhead to the overall performance. After face detection, image $\mathrm{I}_1^{\mathrm{TR}}$ is cropped; the cropped image I_{CR}^{TR} is passed to the feature extraction module. The features of the cropped region are extracted using the ORB descriptor, resulting in a feature vector F_i^{ORB}. This feature vector is transmitted to the cloud where the feature vector F_i^{ORB} is extracted for all the images $\mathrm{I}_{(3N/5)+10}^{\mathrm{TR}}$ in đ$_{\mathrm{TR}}$ and is given to the Bag of Visual Words as input. Bag of Visual Words performs clustering and generates a histogram based on the frequency of the repeated visual words. The output from the Bag of Words is the developed vocabulary which is then used to fuse the feature vectors into 1-D feature vector in the developed vocabulary. After applying these operations on $\mathrm{I}_{(3N/5)+10}^{\mathrm{TR}}$, the feature vector is labeled according to the corresponding label of emotions, $\{L_c = L_1 L_2 ....L_7\}$ where c is the standard defined seven classes of facial expressions. This labeled feature vector is passed to the classifier C which is multi-class SVM for training. In the testing phase, the image $\mathrm{I}_i^{\mathrm{C}}$ is taken from the cross validation set, i.e., testing dataset đ$_{\mathrm{CR}}$, followed by the same procedure for image classification, image enhancement, face alignment, cropping the face region in the input image and then used for the feature extraction as conducted in the training phase. The feature vector F_i^{ORB} extracted for the predictive/testing dataset is cross validated against the label L_c over the train multi-classifier C which is, again, SVM (as it was in the case of training). The whole procedure and its detailed framework are presented in Fig. 2 while the description of parameters is given in Table 1. The overall flow of the proposed method is given in Table 2.

3.1. Image pre-processing

The first step of the proposed framework is to preprocess the image so that the overall quality of the entire image is improved by removing noise. In performing this step, image I is processed prior to actual processing for enhancing its quality. Initially, the image might contain noise or blurring artifacts which can affect the recognition accuracy. Therefore, to remove noise and extract important and relevant information, the Gabor filter is used. This filter has better response to edges and points where texture varies significantly, showing its positive impact on them. Similarly, for preserving important information, we used a median filter, sized 4 × 4, and applied it to image I. In our experiment, when low-contrast or low-resolution images are used then the recognition rate is always degraded. To get rid of any abnormality in lights and background noise of the image, we applied the histogram equalisation technique to enhance the image contrast and also to normalise its illumination effects as shown in Fig. 3.

For further tuning of the Gabor filter, readers are referred to the original paper [20] and documentation of the OpenCV [4].

3.2. Face detection

After the pre-processing when the image quality is enhanced before the face detection step, face detection is the major step in the recognition system because we are concerned with the face only; and all facial landmarks related to expressions exist on the face only. The majority of the face detection algorithms are based on the whole region of the face because

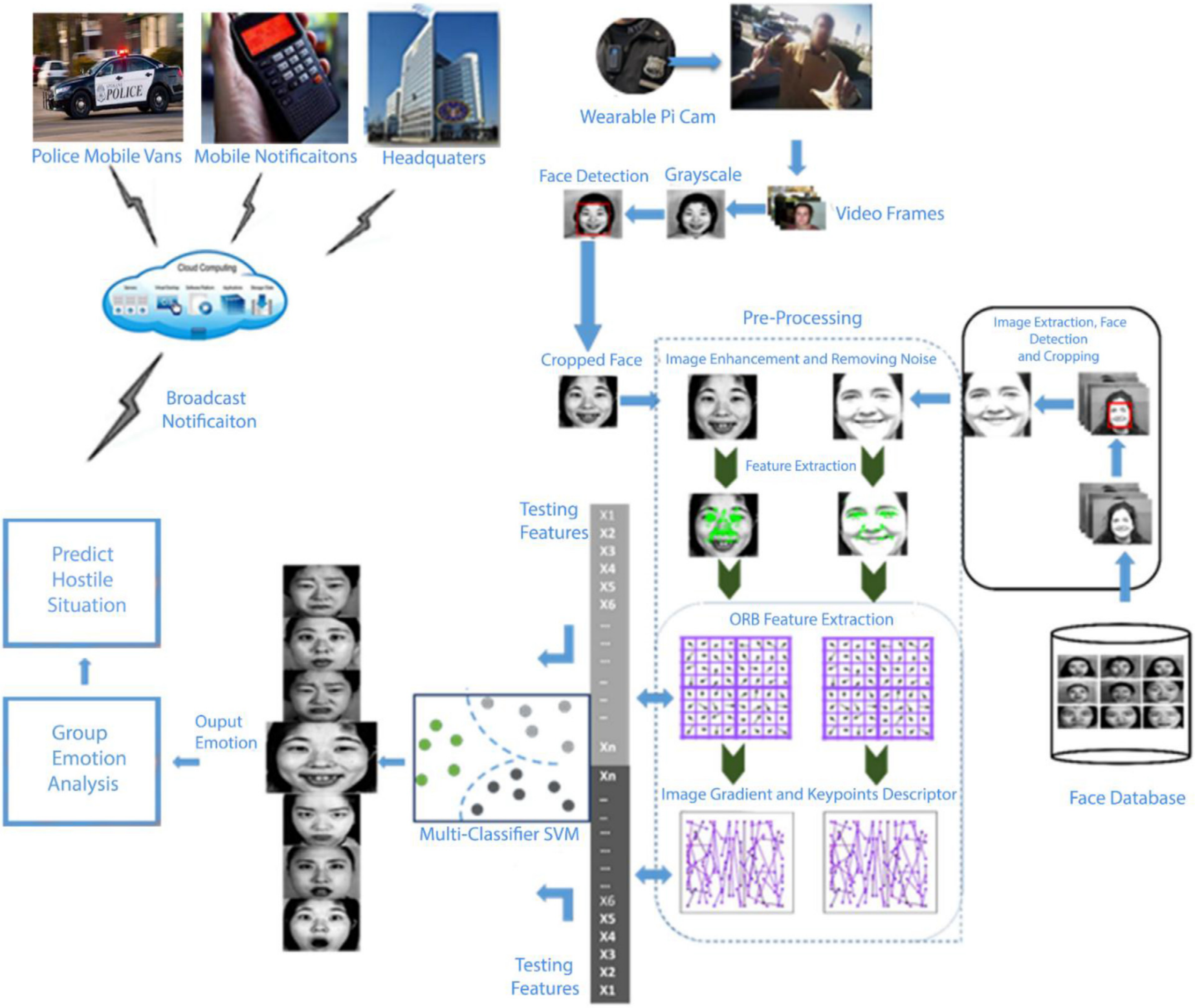


Fig. 2. Facial expression recognition framework using Raspberry Pi.

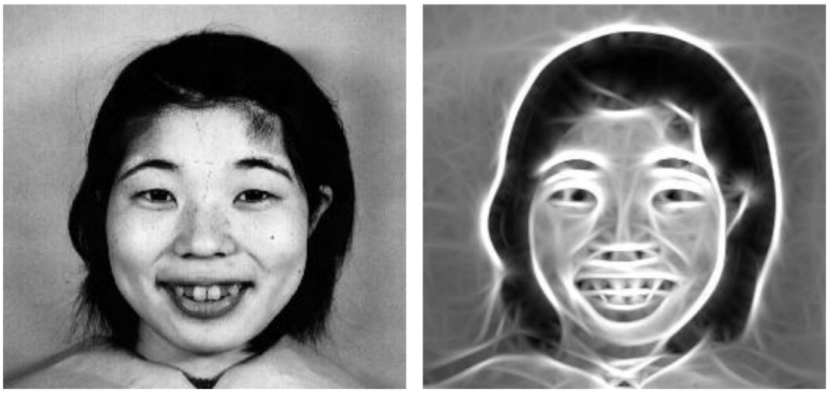


Fig. 3. Showing effect after applying the Gabor filter.

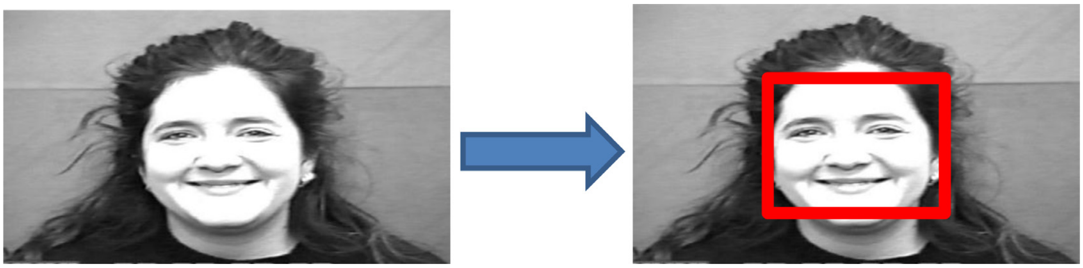
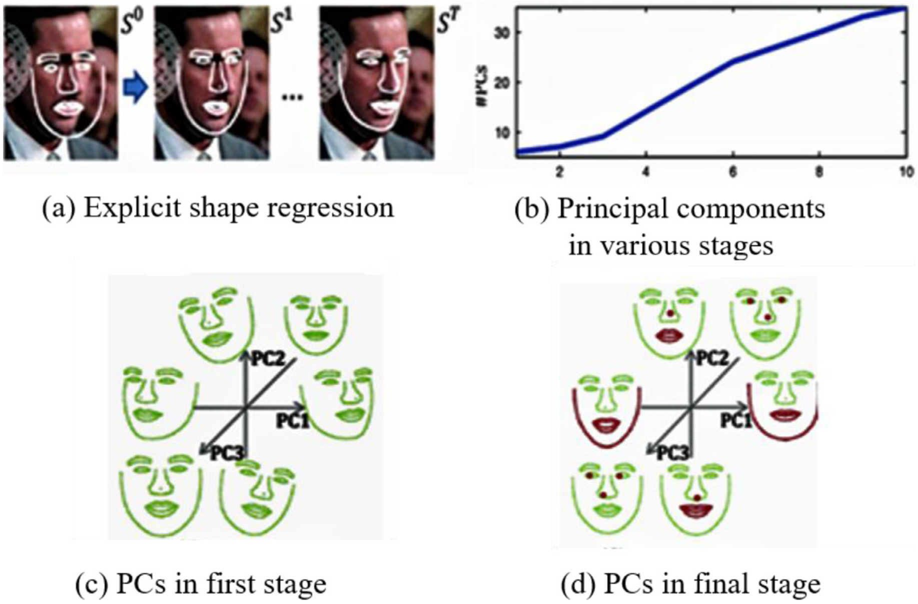


Fig. 4. Simple face detection process using Viola Jones algorithm.



(a) Explicit shape regression

(b) Principal components in various stages

(c) PCs in first stage

(d) PCs in final stage

Fig. 5. Face alignment algorithm using boosted regressor method [17].

sometimes the detection is either inaccurate or too difficult due to movement of the head or with performing different poses.

An automated facial expression recognition (FERs) system often consists of mainly three modules: Face detection, Face features extraction and facial expression recognition. To build an FER system, it is always better to use a robust and efficient algorithm to detect face regardless of head pose, orientation, scale and lighting effects. For this purpose, numerous techniques [36] have been suggested since the first one in 1990. We used the Viola and Jones [44] algorithm in our proposed work for face detection, due to its simplicity, robustness and accuracy, in terms of face detection on Raspberry Pi. It is reliable as compared to state-of-the-art face detection algorithms [39,17], such as deep face, deep dense face detection algorithms, and also due to its resource consumption and time complexity. The Raspberry Pi camera captures the video and extracts the video frames as input image. The image $\boldsymbol{I}$ is cropped and the cropped image $\boldsymbol{I}^{\boldsymbol{F}}$ is converted to gray-scale. The resultant gray-scale image is fed to the features extraction module where features are extracted and are used in the sub-sequent steps. A simple example of face detection is shown in Fig. 4.

3.3. Face alignment

Face alignment refers to identify and fix the geometric structure of the faces in the input frame. This step is usually performed to achieve a good canonical alignment of the face using translation, scale, and rotation. We aligned the face by using the algorithm proposed by Cao et al. [5] He suggested that the face landmarks such as nose, chin, eyes, and mouth are mostly essential for face alignment. A face shape $S=[a1, b1 \ldots aN_{fp}, bN_{fp}]^{T}$ consists of N_{fp} facial landmarks. Given a face image, the aim is to estimate a shape S that is as close as possible to the true shape S^, i.e. Minimising ||S-S^||2. To use the whole face region as input and random ferns as the regressor, shapes of the facial landmarks is expressed as linear combinations of training shapes. The boosted regressor can be used to effectively infer the shape and the early regressor can handle large shape variations with guaranteed accuracy. Therefore, the shape constraint is enforced from coarse to fine face landmarks automatically. The process is described in Fig. 5.

Table 2
Proposed algorithm.

```
Input: Database of face image đ.

I. Divide the database into training $đ_{TR}$ and testing dataset $đ_{CR.}$

II. Training

        For i = 1 to size of ($đ_{TR}$) = $I^{TR}_{(3N/5+10)}$, where N is the total number of images in đ.
        a. Select an image $I^{TR}_1$ from $đ_{TR.}$
        b. Pre-Process the selected $I^{TR}_1$ by applying Gabor filter, Histogram equalisation and median filter:
            $I^E \leftarrow$ Pre-Processing ($I^{TR}_1$)
            c. Detect the face region through viola and Jones algorithm. $I^F \leftarrow$ Viola and Jones ($I^E$).
            d. Crop the face region $I^F \leftarrow$Cropped ($I^F$).
            e. Use boost-regressor face alignment technique to align the face for effective feature extraction.
            f. Extract the ORB features into features vector.
            g. Develop a bag-of-visual-words vocabulary.
            h.Fuse feature vectors into 1-D feature vector.
        End

III. Label all $I^{TR}_{(3N/5)+10}$ feature vectors $F^{ORB}_i$ in the previous step according to the corresponding expression labels L = [Happy, Fear, Disgust, Anger, Neural, Sad, Surprise]

IV. Train the classifier by passing all the feature vectors with their corresponding labels.

V. Conduct cross validation:

        for i = 1 to the size of $đ_{CR} = I^{TR}_{(3N/5)+10}$
        i. Select an image $I^C_i$ from $đ_{CR.}$
        ii. Repeat the sub-steps from b to h of step III (training step) for the images in $đ_{CR.}$
        End

VI. Pass the resultant feature vector of each test image in đCR to the classifier C for cross-validation and updates the classifier accordingly.

Output: Predict L with a respective category L = [Angry, disgust, fear, Happy, Neutral, Sad, Surprise]
```

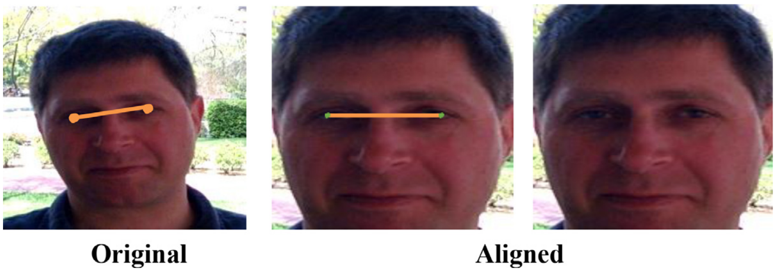


Fig. 6. Unaligned image is aligned after applying the boost regressor algorithm.

Constraints in the shape are preserved and effectively learned in a coarse using a boosted regressor: (**a)** The shape is then adaptively refined by the increments in shape which is learnt by the boosted regressor in different stages; (**b)** Using 87 facial landmarks, the intrinsic dimensions of the face is learnt and is incremented in 10 stage regressor, (**c, d**) are the three principle components of the shape and is incremented in the first and final stage.

The main aim was to transform the image in the output such that face in the entire image is in the center of the resultant image and rotated such that the eyes are on horizontal line. For instance, the face will be rotated such that the eyes lie along the same Y-coordinate and are scaled such that the size of the face is identical. Fig. 6 shows the corrected, aligned version

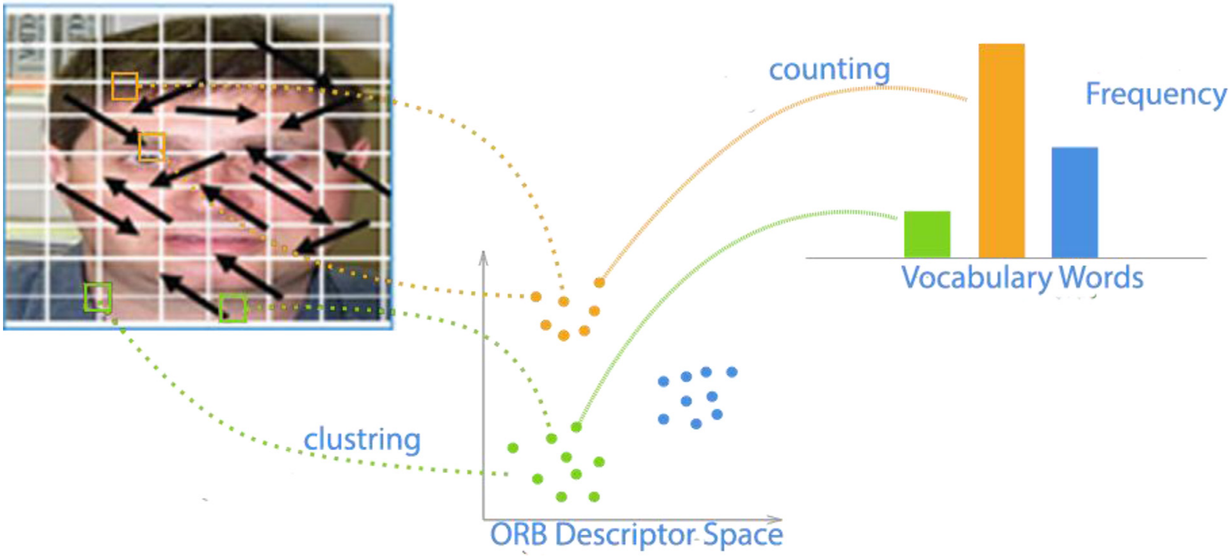


Fig. 7. ORB features extraction using Bag of Visual Words.

of the face where the line according to the whole image is not horizontal in unaligned image while horizontal in the aligned image.

3.4. Feature extraction

Features are important points collected from an image based on certain conditions defined in a feature descriptor. There are several feature extraction methods for feature collection; some use edges while others make use of corners or other statistical or geometrical features. The main challenge of all feature descriptors is that pixels variations can cause significant change in the collection of features. Some feature descriptors are prone to change in orientation and to the introduction of noise. FER is a delicate process and small changes in a few pixels can change the emotions drastically. For this purpose, we chose ORB feature descriptor, which is less prone to change in orientation and performs well under noisy conditions. We also make use of the Bag of feature, also called Bag of Visual Words (BoVW), which is a standard feature descriptor itself. In this method, the local features of training images are extracted ORB. The reason that we use ORB is because it is fast, robust, and uses local feature descriptor for recognising the object, registration, and then classification.

The working mechanism used by ORB with BoVW is similar to SIFT but faster and more robust than SIFT or SURF [29]. Moreover, ORB is less resource-intensive in comparison with SIFT and SURF. Fig. 7 shows the steps involved.

ORB points are extracted and given to the Bag of Visual Words as input. BoVW performs clustering and generate a histogram based on the frequency of the repeated visual words. The output from the Bag of Words is a developed vocabulary that is used for training the SVM classifier.

3.5. Multi-Class SVM based classification

Researchers have presented different techniques for the evaluation of various classifiers for emotion classification. The intensity variation is the main problem in emotion recognition. The classifier must be able to classify these emotions accurately under these variations and constraints. Lucey et al. [25] evaluated the static classifiers for emotion recognition. For instance, the Tree Augmented Naïve Bayes and Decision tree [11] and dynamics classifier, such as single and multi-level Hidden Markov Model [10], are also used for emotion classification. It has been proved experimentally that in the presence of various constraints, the support vector machine (SVM) plays an effective and vital role in multi-classification [43].

From our previous work [32], we investigated and suggested that image classification is a very costly process and takes up lots of resources and the ORB descriptor is more appropriate for smaller, more resource-constrained devices. We use the Multi-class SVM due to its simplicity to implement on Raspberry Pi device. As the resources available on Raspberry Pi is very limited, so classification process take much time and effect the user experience, so its proven that SVM is very efficient, furthermore a thorough investigation is required for the comparison of the performance of different variants of SVM on Raspberry Pi.

Due to all these aspects, we evaluated the multi-class SVM as a classifier in our research work for emotion recognition and classification. In this work, the classification is based on more than two classes, thus we are using multi-class SVM, which classify the output into seven classes.

Table 3
System configuration.

Name	Configuration
Imaging libraries	OpenCV 3.1.0, Scikit-Learn, Scikit-Image, MATLAB Support Package for Raspberry Pi
Libraries	Numpy, SciPy, PyLab, Matplotlib, RPI.GPIO
Programming language	Python 2.7
Operating system	Noobs (Raspbian)

Initially, SVM was mainly proposed for binary classification but later due to variety of methods it tuned into Multi-class classification. Two methods are mostly used for multi-classification problems: 1) by reducing it to multiple binary classification problems 2) and one verses one. In our proposed system, the one versus rest strategy is evaluated.

4. Experimental results

The detail evaluation of the proposed work is provided in this section on various standard databases. Next, an extensive comparative analysis is performed with state-of-the-art emotion recognition techniques using both the quantitative and qualitative assessments. We used three standard databases where each database is divided into training and testing set. The proposed system is divided into 70% of training and 30% of testing images, which is the standard dividing scale for training. By varying training and testing sets of images, changing the features, and varying databases, we performed different experiments. The whole platform used for this purpose was Python using OpenCV and Raspberry Pi to acquire the stream from the camera. The Raspberry Pi device is used with the following specification. This device has a Broadcom BCM2837 system on a chip. Besides this, it has an SD card slot, video Core IV GPU, 1.2 GHz processor and ARM Cortex-A53. Its GPU is capable of video playback using H.264 which can play Blu-ray quality videos at 40 MB/s. The proposed work is done on the system with the specifications given in Table 3.

4.1. Dataset

The proposed framework has been tested on many datasets available publically. Each dataset has its own advantages and disadvantages. The number of samples is also different in each dataset. The detailed explanation of each dataset is given below.

4.1.1. Cohn Kanade

CK+ is a publically available dataset for evaluation of the facial expression analysis systems [25]. The dataset consists of mainly continuous shots of images, which is the content of both the emotions and non-emotions images, i.e., the first image of the series is a neutral emotion, while the last image shows the final emotion. The dataset consists of facial expressions of 123 individuals and seven basic emotions are performed by each individual. Fig. 8(a) illustrates various sample images collected from this dataset, showing different emotions starting from neutral ending with surprise emotion. In this dataset, the same structure is presented where the expressions are from neutral to a certain emotions. We used about 578 images from this dataset in our proposed work.

4.1.2. MMI facial expression database

This dataset [31] is collected from more than 20 individuals containing both genders (45% female and 55% male), belonging to different continents like South America, Asia, and Europe. Both male and female have performed 80 sequences of expressions with six basic emotions Happy, Sad, Disgust, Surprise, Angry, Fear and Neutral. These images are taken from a sequence of videos which contain all the emotions including posed and non-posed. As these are video frames, so the timestamp of the video is hard-embedded on the images, which is removed during the cropping phase. After face detection, a total of 270 frames were extracted from the videos. Fig. 8(b) illustrates some examples from the MMI database.

4.1.3. Japanese female facial expression (JAFFE) database

This dataset [26] consists of 213 images, which are taken from 10 Japanese females. Each subject has shown the same set of seven basic emotions discussed earlier. All 213 images are used for experiments in this work. Few images from this dataset are shown in Fig. 8(c), illustrating few emotions by JAFFE individuals.

4.1.4. Yale faces

This dataset is collected from 15 individuals, having 165 grayscale images, hence 11 images per individual. Each individual has performed different emotions and configuration: wink, surprised, normal, sad, happy, sleepy, with glasses, no glasses, left-light, right-light, and center-light. In our experiments, we collected the emotions with no glasses and light effects. In the given Fig. 8(d), we gave different emotions of different individuals because there was no such individual who had performed all the emotions. However, we did not perform experiments on this dataset because there were less amount of images

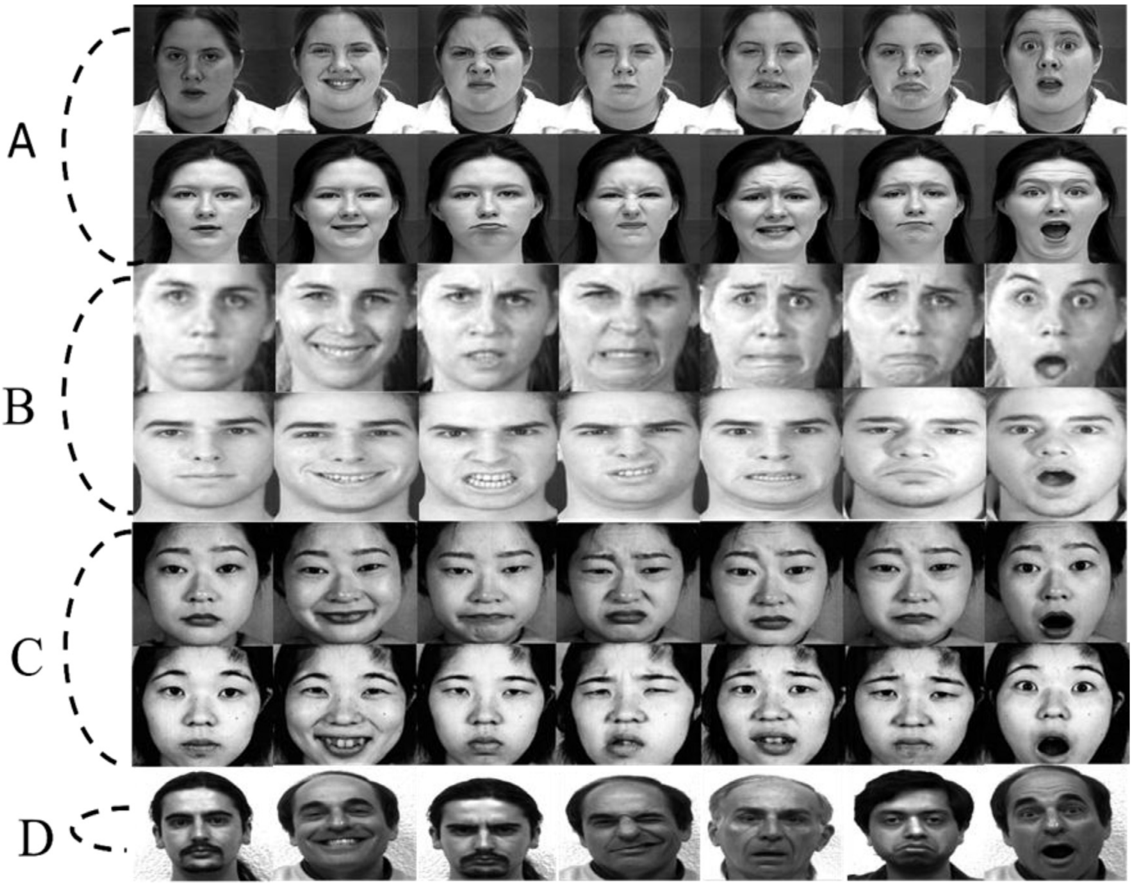


Fig. 8. (a) The extended Cohn-Kanade (b) MMI (c) JAFFE (d) Yale Faces.

Table 4
Number of images in each dataset used to evaluate the proposed framework.

Emotions	Cohn–Kande (CK+)	MMI	JAFFE	Yale Faces	Total
Angry	61	35	30	0	128
Disgust	88	34	29	0	151
Fear	25	35	31	0	91
Happy	105	34	30	15	179
Neutral	122	28	30	15	190
Sad	58	34	30	15	128
Surprise	119	34	30	15	193
Total	578	234	210	60	1082

regarding the emotions, so we just studied it and used a portion of the dataset for training the classifier to make it more accurate.

On the above mentioned datasets, we performed our experiments – except the Yale faces dataset which is the collection of rare amount of all emotions and it does not fulfill our needs to perform the experiments. We perform the experiments with different parameters to achieve the highest possible accuracy without compromising the user experience. All the experiments with results are given as follows:

The proposed framework is evaluated using different publically available datasets. The number and quality of images in each dataset is different and hence it affects the overall performance of the system. The total number of images in each dataset is given in Table 4.

Table 5 shows that the overall accuracy of the SVM classifier using a 70/30 split of training and testing is 94% using ORB features. The cluster size in the bag-of-words is 700 and the cross validation value is 0.7. Fear is identified at 100% success rate. It is because of the less number of images available in the dataset. Realistically the level of fear cannot be measured with 100% accuracy using only expressions. It requires more parameters such as perspiration and heart rate.

Table 5
Confusion matrix of the proposed method on MMI training (70%) and testing (30%) using ORB points, cluster size of 700 and the cross validation value of 0.7.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	78.57	0	0	0	0	21.43	0
Disgust	0	84.62	7.69	7.69	0	0	0
Fear	0	0	100	0	0	7.14	0
Happy	0	0	0	92.31	0	0	0
Neutral	0	0	0	0	100	0	0
Sad	0	0	0	0	0	100	0
Surprise	0	0	0	0	0	7.69	92.31
Average Accuracy	**94%**						

Table 6
Confusion matrix of the proposed method on MMI training (70%) and testing (30%) using SIFT points, cluster size of 500 and cross validation value of 0.5.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	78.57	0	0	0	7.14	14.29	0
Disgust	0	92.31	0	0	0	7.69	0
Fear	0	14.29	85.71	0	0	0	0
Happy	0	0	0	92.31	7.69	0	0
Neutral	0	0	0	0	100	0	0
Sad	0	0	0	0	0	100	0
Surprise	0	0	0	0	7.69	0	92.31
Average Accuracy	**93%**						

Table 7
Confusion matrix of the proposed method on MMI training (70%) and testing (30%) using SURF points, cluster size of 700 and cross validation value of 0.5.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	100	0	0	0	0	0	0
Disgust	0	100	0	0	0	0	0
Fear	0	0	100	0	0	0	0
Happy	0	0	0	100	0	0	0
Neutral	0	0	0	0	100	0	0
Sad	0	0	0	0	0	100	0
Surprise	0	0	7.69	0	0	0	92.31
Average Accuracy	**99%**						

Table 8
Confusion matrix of the proposed method on JAFEE training (70%) and testing (30%) using SURF points, cluster size of 250 and cross validation value of 0.3.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	88.89	0	0	0	0	11.11	0
Disgust	0	87.50	0	0	0	0	12.50
Fear	10	0	80	0	0	0	10
Happy	0	0	0	77.78	0	11.11	11.11
Neutral	0	0	0	0	88.89	0	11.11
Sad	0	11.11	0	0	22.22	55.56	11.11
Surprise	0	0	22.22	0	0	0	77.78
Average Accuracy	**81%**						

The highest level of accuracy that we achieved with MMI dataset and SIFT points was with the following setting: size of K: 500 and cross validation: 0.5. The points collected by SIFT were very good for classifying the emotions, but the memory intake and time consumption on Raspberry Pi was poor. Confusion matrix of the SIFT points is presented in Table 6.

Table 7 shows the accuracy of the SURF points for emotions recognition. It is by far the most accurate prediction achieved in any of our experiments. The size of the cluster was constant at 700 while the cross validation value was fixed to 0.5. Any more than this value resulted in reduction of the accuracy percentage. Surprise emotion was difficult to predict for all datasets and it was often confused with normal or neutral.

The second dataset that we used for the purpose of experimentation is JAFEE. The details of this dataset were already presented in the above section. The number of images in JAFEE dataset is less than MMI and hence the classifier did not achieve much accuracy. With SURF points and with cross validation value of 0.3, the classifier reached an accuracy of 81% as shown in Table 8.

Table 9
Confusion matrix of the proposed method on JAFEE training (70%) and testing (30%) using SIFT points, cluster size of 500 and cross validation value of 0.7.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	77.78	0	0	0	0	22.22	0
Disgust	0	75.00	0	0	0	25.00	0
Fear	0	0	80	0	10	10	0
Happy	0	0	0	88.89	11.11	0	0
Neutral	0	0	0	0	100	0	0
Sad	0	11.11	0	0	11.11	77.78	0
Surprise	0	0	11.11	0	11.11	0	77.78
Average Accuracy	**86%**						

Table 10
Confusion matrix of the proposed method on JAFEE training (70%) and testing (30%) using ORB points, cluster size of 700 and cross validation value of 0.5.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	66.67	11.11	0	0	0	22.22	0
Disgust	12.50	75.00	0	0	12.50	25.00	0
Fear	0	0	80	0	20	10	10
Happy	0	0	0	77.78	22.22	0	0
Neutral	0	0	0	0	100	0	0
Sad	11.11	0	0	0	22.22	66.67	0
Surprise	0	0	0	0	44.44	0	55.56
Average Accuracy	**78%**						

Table 11
Confusion matrix of the proposed method on CK+ training (70%) and testing (30%) using ORB points, cluster size of 700 and cross validation value of 0.3.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	0	11.11	0	5.56	77.78	0	5.56
Disgust	0	34.62	0	3.85	50	0	11.54
Fear	0	14.29	0	42.86	42.86	0	0
Happy	0	0	0	53.33	36.67	0	10
Neutral	0	0	0	2.78	80.56	0	16.67
Sad	0	5.88	0	0	82.35	5.88	5.88
Surprise	0	0	0	0	30.56	0	69.44
Average Accuracy	**53%**						

Table 12
Confusion matrix of the proposed method on CK+ training (70%) and testing (30%) using SIFT points, cluster size of 500 and cross validation value of 0.3.

Emotions	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	0	11.11	0	5.56	77.78	0	5.56
Disgust	0	34.62	0	3.85	50	0	11.54
Fear	0	14.29	0	42.86	42.86	0	0
Happy	0	0	0	53.33	36.67	0	10
Neutral	0	0	0	2.78	80.56	0	16.67
Sad	0	5.88	0	0	82.35	5.88	5.88
Surprise	0	0	0	0	30.56	0	69.44
Average Accuracy	**53%**						

The ORB points collected with cross validation of 0.5 and K size of 700 resulted in the highest accuracy of 78% as shown in Table 10. This is not as accurate as SIFT or SURF. But the running time of ORB is the smallest, making it ideal for small devices with constraints such as Pi. We present the results of all three as one may prefer accuracy over the time as shown in Table 9 and Table 11.

Similarly, the CK+ dataset has higher number of images, but the noise and quality of the dataset make it difficult to train an accurate model. The CK+ dataset has several disadvantages. Almost all the happy emotions can also be classified as neutral, making it difficult even for humans. Human emotions are a mix of multiple emotions and assigning one class is not practical; but in quantifiable terms, the accuracy achieved using CK+ dataset was less than MMI and JAFEE.

Table 12 shows the highest accuracy achieved by SIFT points was 53% with the cross validation value of 0.3.The proposed system is compared with several state-of-the-art techniques including [33,9,6,19]. The resultant accuracy and efficiency of the proposed method is compared with the output result and accuracy of the different benchmarked approaches. Due to

Table 13
Comparison of the proposed method with state-of-the-art-methods.

References	Databases	Feature Extraction	Classifiers	Average Accuracy
[33]	MMI	HOG + U-LTP	SVM	98.20
	JAFFE			95.71
	CK+			99.68
[9]	JAFFE	$LPQ_{+es-}LBP$	SVM	94.88
[6]	CK+	HOG	SVM	98.8
[19]	JAFFE	Radial Encoding of Local Gabor features	KNN	89.67
Proposed method	MMI	ORBSURFSIFT	SVM	99.1
	CK++			90
	JAFFE			92

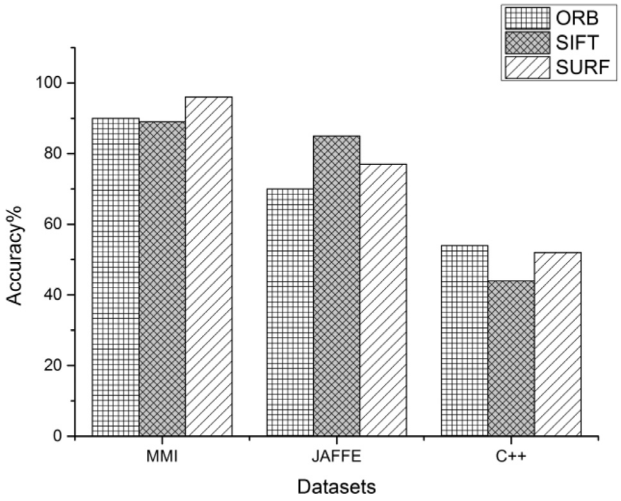


Fig. 9. Accuracy measurements of different datasets used.

the production of state-of-the-art performance using similar testing technique on the same dataset, we selected these approaches. It has been shown in the Table 13 that the proposed system outperforms others existing methods using the same datasets. The accurate recognition rate is 90% on MMI dataset and 94% on JAFFE dataset and 90% on the CK+ dataset using variation of cross validation value and no of clusters. Table 13 provides the overall performance and their comparison among the proposed method and the existing systems by using the same dataset MMI, CK+ and JAFFE datasets.

The experiments on all datasets suggest that the SVM classifier worked significantly well on SURF points using MMI dataset as shown in Fig. 9. Using the SURF points on almost all datasets, the performance of the classifier is much better than either of SIFT or ORB.

Fig. 10 shows emotions recognised in all datasets with varying results. The happy emotion was recognised easily, as it does not conflict with any other emotions. However, fear and disgust remained the lowest in prediction accuracy as they are both difficult to predict even for humans, and the number of images in these emotions were less comparatively. Furthermore, the neutral and surprised were often confused by the classifier as the level of surprise is not common in every person.

The proposed system has also been tested on group images and can predict the collective emotions of the group as shown in Fig. 11. Although the technique is tested only on images and not videos, yet it can be applied on live video streams with certain video frame sampling. However, the sheer resources required for such computation is very resource intensive. Raspberry Pi, being a small device with limited resources cannot process huge videos in real-time and it will require cloud-services for group emotion classifications. In future work, we aim to propose a similar platform that can perform group emotion analysis in real-time.

5. Conclusion

Real-time suspicious activity detection and recognition without interaction with the subject is a challenging task. Facial expression recognition technology and the embedded processing capabilities of portal devices such as Raspberry Pi can significantly help law enforcement agencies in identification of suspicious activities in a cost-effective way, providing smart security. Considering this motivation, we have proposed an efficient suspicious activity detection framework using cloud services and Raspberry Pi. Raspberry Pi not only provides expedient platform for detection and feature extraction, but also reliably communicates with the cloud for assistance where necessary. The proposed framework automatically extracts the face

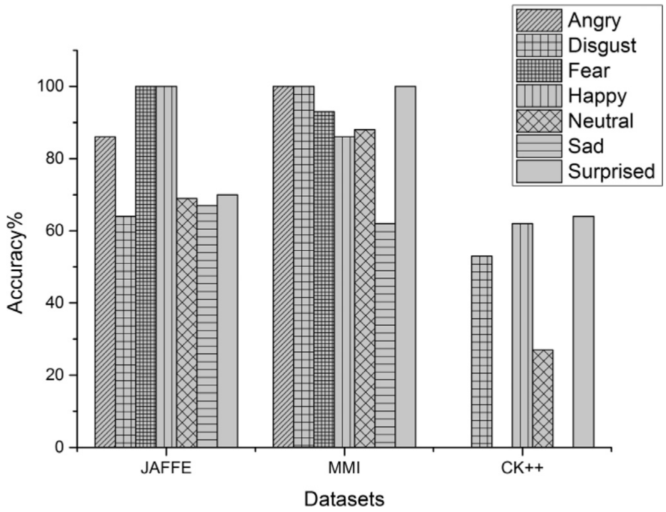


Fig. 10. Facial Expression Recognition accuracy of the all datasets.

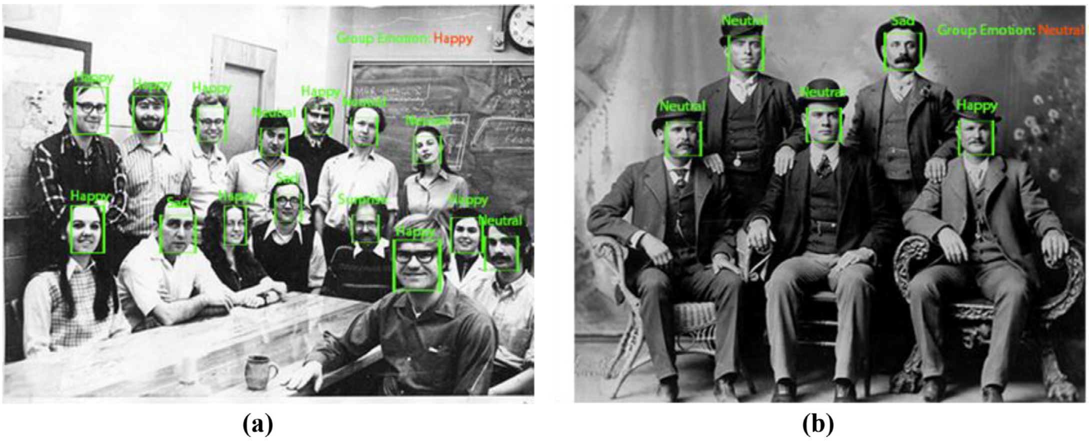


Fig. 11. (a) Image obtained from Rossmann's Lab, Purdue University, USA (b) Image obtained from Butch Cassidy and the Wild Bunch Butch Cassidy (Robert Leroy Parker) 1866.

using Viola Jones algorithm, followed by its features extraction using ORB algorithm, which is significantly better than SURF and SIFT as proved in our experiments. The experimental evaluation suggests that ORB has significant advances over SIFT and SURF as it is more robust and is ideal for processors with limited capabilities. Quantitative and qualitative evaluation using three different datasets validate the effectiveness of the proposed framework for smart security in law-enforcement services in smart cities.

Our future work will focus on investigating other action and event detection and analysis methods [7,8,41], convolutional neural networks and deep belief networks with their performance on Raspberry Pi. Further, the face detection technology [32] with encryption [28] can be combined for suspicious activity detection and further improving the security in smart cities.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (no. 2016R1A2B4011712).

References

[1] M. Al-Shabi, W.P. Cheah, T. Connie, Facial expression recognition using a hybrid CNN-SIFT aggregator, arXiv preprint arXiv:1608.02833, (2016).
[2] B. Allaert, J. Mennesson, I.M. Bilasco, C. Djeraba, Impact of the face registration techniques on facial expressions recognition, Signal Process. Image Commun. 61 (2018) 44–53.
[3] A.B. Ashraf, S. Lucey, J.F. Cohn, T. Chen, Z. Ambadar, K.M. Prkachin, P.E. Solomon, The painful face–pain expression recognition using active appearance models, Image Vision Comput. 27 (2009) 1788–1796.

[4] G. Bradski, A. Kaehler, Learning OpenCV: Computer vision With the OpenCV Library, O'Reilly Media, Inc., 2008.
[5] X. Cao, Y. Wei, F. Wen, J. Sun, Face alignment by explicit shape regression, Int. J. Comput. Vision 107 (2014) 177–190.
[6] P. Carcagnì, M. Coco, M. Leo, C. Distante, Facial expression recognition and histograms of oriented gradients: a comprehensive study, SpringerPlus 4 (2015) 645.
[7] X. Chang, Z. Ma, Y. Yang, Z. Zeng, A.G. Hauptmann, Bi-level semantic representation analysis for multimedia event detection, IEEE Trans. Cybern. 47 (2017) 1180–1197.
[8] X. Chang, Y.-L. Yu, Y. Yang, E.P. Xing, Semantic pooling for complex event analysis in untrimmed videos, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 1617–1632.
[9] W.-L. Chao, J.-J. Ding, J.-Z. Liu, Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection, Signal Process. 117 (2015) 1–10.
[10] I. Cohen, N. Sebe, A. Garg, L.S. Chen, T.S. Huang, Facial expression recognition from video sequences: temporal and static modeling, Comput. Vis. Image Underst. 91 (2003) 160–187.
[11] I. Cohen, N. Sebe, F. Gozman, M.C. Cirelo, T.S. Huang, Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data, in: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on,, IEEE, 2003 I-I.
[12] P.M. Corcoran, C. Iancu, Automatic face recognition system for hidden markov model techniques, New Approaches to Characterization and Recognition of Faces, InTech, 2011.
[13] S. Datta, D. Sen, R. Balasubramanian, Integrating geometric and textural features for facial emotion classification using SVM frameworks, in: Proceedings of International Conference on Computer Vision and Image Processing, Springer, 2017, pp. 619–628.
[14] J. Edwards, H.J. Jackson, P.E. Pattison, Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review, Clin. Psychol. Rev. 22 (2002) 789–832.
[15] P. Ekman, E.L. Rosenberg, What the Face reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), Oxford University Press, USA, 1997.
[16] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognit. 36 (2003) 259–275.
[17] S.Z. Gilani, A. Mian, P. Eastwood, Deep, dense and accurate 3D face correspondence for generating population specific deformable models, Pattern Recognit. 69 (2017) 238–250.
[18] S.J. Goyal, A.K. Upadhyay, R. Jadon, R. Goyal, Real-life facial expression recognition systems: a review, in: Smart Computing and Informatics, Springer, 2018, pp. 311–331.
[19] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, H. Lin, Facial expression recognition using radial encoding of local Gabor features and classifier synthesis, Pattern Recognit. 45 (2012) 80–91.
[20] J.P. Jones, L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, J. Neurophysiol. 58 (1987) 1233–1258.
[21] L. Juan, O. Gwun, A comparison of sift, pca-sift and surf, Int. J. Image Process. (IJIP) 3 (2009) 143–152.
[22] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition—a review, Comput. Vision Image Underst. 97 (2005) 103–135.
[23] M.S.S. Kumar, R. Swami, An improved face recognition technique based on modular LPCA approach, (2011).
[24] Y.-H. Lee, W. Han, Y. Kim, Emotional recognition from facial expression analysis using bezier curve fitting, in: Network-Based Information Systems (NBiS), 2013 16th International Conference on, IEEE, 2013, pp. 250–254.
[25] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, IEEE, 2010, pp. 94–101.
[26] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with gabor wavelets, in: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, IEEE, 1998, pp. 200–205.
[27] H.Z.A.C.B. Michael, M.J. Malik, SVMKNN: discriminative nearest neighbor classification for visual category recognition, in: Computer Science Division, EECS Department Univ. of California, Berkeley, CA, 2007, p. 94720.
[28] K. Muhammad, R. Hamza, J. Ahmad, J. Lloret, H.H.G. Wang, S.W. Baik, Secure surveillance framework for IoT systems using probabilistic image encryption, IEEE Transactions on Industrial Informatics, 2018, doi:10.1109/TII.2018.2791944.
[29] P. Panchal, S. Panchal, S. Shah, A comparison of SIFT and SURF, Int. J. Innov. Res. Comp. Commun. Eng. 1 (2013) 323–327.
[30] M. Pantic, L.J.M. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1424–1445.
[31] M. Pantic, M. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, in: Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, IEEE, 2005, p. 5.
[32] M. Sajjad, M. Nasir, K. Muhammad, S. Khan, Z. Jan, A.K. Sangaiah, M. Elhoseny, S.W. Baik, Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities, Futur. Gener. Comput. Syst. (2017).
[33] M. Sajjad, A. Shah, Z. Jan, S.I. Shah, S.W. Baik, I. Mehmood, Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery, Clust. Comput. (2017) 1–19, doi:10.1016/j.future.2017.11.013.
[34] M.B. Salter, Passports, mobility, and security: how smart can the border be? Int. Stud. Perspect. 5 (2004) 71–91.
[35] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: integrated recognition, localization and detection using convolutional networks, arXiv preprint arXiv:1312.6229, (2013).
[36] M.I. Sezan, A peak detection algorithm and its application to histogram-based image data reduction, Comput. Vis. Graphics Image Process. 49 (1990) 36–51.
[37] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vis. Comput. 27 (2009) 803–816.
[38] M. Suwa, A preliminary note on pattern recognition of facial emotional expression, The 4th International Joint Conferences on Pattern Recognition, 1978, 1978.
[39] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Closing the gap to human-level performance in face verification. deepface, IEEE Computer Vision and Pattern Recognition (CVPR), 2014.
[40] A. Uçar, Y. Demir, C. Güzeliş, A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering, Neural Comput. Appl. 27 (2016) 131–142.
[41] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, S.W. Baik, Action recognition in video sequences using deep Bi-directional LSTM with CNN features, IEEE Access 6 (2018) 1155–1166.
[42] M.F. Valstar, B. Jiang, M. Mehu, M. Pantic, K. Scherer, The first facial expression recognition and analysis challenge, in: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, 2011, pp. 921–926.
[43] M.F. Valstar, I. Patras, M. Pantic, Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data, Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, IEEE, 2005 76-76.
[44] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, IEEE, 2001 I-I.
[45] G. Yang, Y. Lin, P. Bhattacharya, A driver fatigue recognition model based on information fusion and dynamic Bayesian network, Inf. Sci. 180 (2010) 1942–1954.
[46] J. Yang, D. Zhang, A.F. Frangi, J.-y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 131–137.

[47] M. Yang, L. Zhang, S.C.-K. Shiu, D. Zhang, Monogenic binary coding: an efficient local feature extraction approach to face recognition, IEEE Trans. Inf. Forensics Secur. 7 (2012) 1738–1751.
[48] S. Yang, G. Bebis, Y. Chu, L. Zhao, Effective face recognition using bag of features with additive kernels, J. Electron. Imaging 25 (2016) 013025.
[49] Z. Zhang, F. Li, M. Zhao, L. Zhang, S. Yan, Robust neighborhood preserving projection by nuclear/l2, 1-norm regularization for image feature extraction, IEEE Trans. Image Process. 26 (2017) 1607–1622.
[50] C. Zhu, Y. Zheng, K. Luu, M. Savvides, CMS-RCNN: contextual multi-scale region-based CNN for unconstrained face detection, in: Deep Learning for Biometrics, Springer, 2017, pp. 57–79.