Saliency Based Framework for Facial Expression Recognition

Article *in* Frontiers of Computer Science (electronic) · March 2017

DOI: 10.1007/s11704-017-6114-9

4 authors:

Rizwan Ahmed Khan
University of Lyon
18 PUBLICATIONS **91** CITATIONS
SEE PROFILE

Alexandre Meyer
Claude Bernard University Lyon 1
37 PUBLICATIONS **549** CITATIONS
SEE PROFILE

Hubert Konik
Université Jean Monnet
51 PUBLICATIONS **288** CITATIONS
SEE PROFILE

Saida Bouakaz
Claude Bernard University Lyon 1
71 PUBLICATIONS **317** CITATIONS
SEE PROFILE

Frontiers of Computer Science CN 10-1014/TP
DOI 10.1007/s11704-017-6114-9 ISSN 2095-2228
RESEARCH ARTICLE

Saliency Based Framework for Facial Expression Recognition

Rizwan Ahmed KHAN[1,3], Alexandre MEYER[1], Hubert KONIK[2], Saida BOUAKAZ[1]

1 Université de Lyon, CNRS Université Lyon 1, LIRIS, UMR5205, F-69622, France
2 Université Jean Monnet, Laboratoire Hubert Curien, UMR5516, 42000 Saint-Etienne, France
3 Barrett Hodgson University, Faculty of IT & Computer Sciences, Pakistan

Front. Comput. Sci., **Just Accepted Manuscript** • 10.1007/s11704-017-6114-9
http://journal.hep.com.cn on March 24, 2017

Front.Comput.Sci.
DOI

RESEARCH ARTICLE

Saliency based framework for facial expression recognition

Rizwan Ahmed KHAN (✉)[1,3], Alexandre MEYER (✉)[1], Hubert KONIK [2], Saida BOUAKAZ [1]

[1] Université de Lyon, CNRS
Université Lyon 1, LIRIS, UMR5205, F-69622, France
[2] Université Jean Monnet, Laboratoire Hubert Curien, UMR5516, 42000 Saint-Etienne, France
[3] Barrett Hodgson University, Faculty of IT & Computer Sciences, Pakistan

Abstract This article proposes a novel framework for the recognition of six universal facial expressions. The framework is based on three set of features extracted from the face image: entropy, brightness and local binary pattern. First, saliency maps are obtained by state-of-the-art saliency detection algorithm i.e. "frequency-tuned salient region detection". The idea is to use saliency maps to find appropriate weights or values for extracted features (i.e. brightness and entropy). To validate the performance of saliency detection algorithm against human visual system, we have performed a visual experiment. Eye movements of 15 subjects were recorded with an eye-tracker in free viewing conditions as they watch a collection of 54 videos selected from Cohn-Kanade facial expression database. Results of the visual experiment provided the evidence that obtained saliency maps conforms well with human fixations data. Finally, evidence of the proposed framework's performance is exhibited through satisfactory classification results on Cohn-Kanade database, FG-NET FEED database and Dartmouth database of children's faces.

Keywords Facial expression recognition, classification, salient regions, entropy, brightness, local binary pattern

1 Introduction

Communication in any form i.e. verbal or non-verbal is vital to complete various daily routine tasks and plays a significant role in life. Facial expression is the most effective form of non-verbal communication and it provides a clue about emotional state, mindset and intention [1–4]. Facial expression not only can change the flow of conversation [5] but also provides the listeners a way to communicate a wealth of information to the speaker without even uttering a single word [6]. According to [7, 8] when the facial expressions do not coincide with the other communication i.e. spoken words, then the information conveyed by the face gets more weight in decoding information.

In this article we propose a novel framework that efficiently recognizes six universal facial expressions [9]. These six facial expressions are anger, disgust, fear, happiness, sadness and surprise and are labeled as "universal" as these expressions are proved to be consistent across cultures. To recognize these expressions in real-time it is desirable to reduce computational complexity of feature extraction. In the proposed method, reduction in computational complexity for feature extraction is achieved by processing only some regions ("salient regions") of the face that contain discriminative information. In order to determine which facial region(s) are salient or contains discriminative information we have taken the help of human visual system (HVS). We conducted a visual experiment with the help of an eye-tracker and recorded the fixations and saccades of human observers as they watch the collection of videos showing six universal facial expressions. It is known that eye gathers most of the information during fixations [10, 11]. Eye fixations describe the way in which visual attention is directed towards salient regions in a given stimulus. So, the concept of using an eye-tracking system and recording fixations from human

Received month dd, yyyy; accepted month dd, yyyy

E-mail: {*Rizwan − Ahmed.Khan, Alexandre.Meyer*}*liris.cnrs.fr*

observers is to find that which component(s) of face i.e. eyes, nose, forehead or mouth is important or salient for human observer for a particular expression.

In the proposed framework we used state-of-the-art saliency detection algorithm i.e. *frequency-tuned salient region detection* [12] to algorithmically highlight saliency of different facial regions (see section 1 for details). Then obtained saliency maps were processed with Viola-Jones object detection algorithm [13] to localize only those facial regions that as salient as per HVS. Then localized salient regions were processed to extract features (refer Algorithm 1 for details).

There exist different methods for automatic recognition of universal expressions [14–24] including expression of pain [25]. But to the best of our knowledge none of the proposed method was tested on stimuli containing children faces, except [26]. Secondly, we have found one shortcoming in all of the reviewed methods for automatic facial expression recognition (see Section 2 for literature review) that none of them try to mimic human visual system in recognizing them. Rather all of the methods, spend computational time on whole face image or divides the facial image based on some mathematical or geometrical heuristic for features extraction. We argue that the task of expression analysis and recognition could be done in more conducive manner, if only some regions are selected for further processing (i.e. salient regions) as it happens in human visual system. Thus, our contribution in this paper are as follows:

a. We have statistically determined which facial region(s) is salient according to human vision for a particular expression by conducting a psycho-visual experiment. The experiment has been carried out using eye-tracker which records the fixations and saccades of human observers as they watch the collection of videos showing six universal facial expressions. Salient facial regions for specific expressions have been determined through the analysis of the fixation data.
b. We have validated the classical results from the domain of human psychology, cognition and perception by a novel approach which incorporates eye-tracker in the experimental methodology protocol. At the same time results have been extended to include all the six universal facial expressions which was not the case in the classical studies.
c. We show that reasonable facial expression recognition accuracy is achievable by using proposed framework. Our proposed framework creates novel feature space by extracting multi-resolution local binary pattern (LBP) [27], brightness and entropy features only from the salient facial region(s). The novel framework for facial expression recognition achieved recognition accuracy in the range of 85% to 95% on extended Cohn-Kanade (CK+) database [28] and FG-NET FEED [29] database. The framework could be used for real-time applications as its current implementation (Matlab 7.6 on Windows PC with 1.8GHz processor and 1GB of RAM) can process 4 frames per second which is enough since expressions do not change abruptly.
d. We have tested our proposed framework on stimuli containing children's faces i.e. Dartmouth database of children's faces [30] and achieved average facial expression recognition accuracy of 82.3%. Usually children show expressions in a subtle way, which creates large inter and intra population variations in stimuli as opposed to adults. Thus, the problem of facial affect analysis for children is tricky.

Rest of the article is organized as follows: literature review is presented in Section 2. All the details related to visual experiment, results and analysis of eye-tracking data is presented in Section 3. Section 4 presents the novel framework for automatic facial expression recognition. Section 5 presents results of proposed framework on databases containing adult and children faces. This is followed by conclusion.

2 Literature review

Generally, facial expression recognition (FER) system consist of three steps: face detection, feature extraction and expression classification. Feature selection is one of the most important step to successfully analyze and recognize facial expressions automatically. The optimal features should minimize within-class variations of expressions while maximize between class variations. In literature various methods are employed to extract facial features and these methods can be categorized either as appearance-based methods or geometric feature-based methods where the shapes and locations of facial components are extracted to form a feature vector [31].

Appearance-based methods. One of the widely studied methods to extract appearance information is based on Gabor wavelets [14, 15]. Littlewort et al. [14] has shown a

high recognition accuracy (93.3% for Cohn-Kanade facial expression database [32]) using Gabor features. They proposed to extract Gabor features from the whole face and then selected the subset of those features using AdaBoost method. Tian [15] has used Gabor wavelets of multi-scale and multi-orientation at the “difference” images. The difference images were obtained by subtracting a neutral expression frame from the rest of the frames of the sequence. Generally, the drawback of using Gabor filters is that it produces extremely large number of features and it is both time and memory intensive to convolve face images with a bank of Gabor filters to extract multi-scale and multi-orientational coefficients. Another promising approach to extract appearance information is by using Haar-like features. Yang et al. [16] extracted Haar-like features from the facial image patches (49 sub-windows). Compositional features based on minimum error based optimization strategy were build within the Boosting learning framework. The proposed method was tested on Cohn-Kanade facial expression database [32] and it achieved average recognition accuracy of 92.3% on the apex data (last three frames of the sequence) and 80% on the extended data(frames from onset to apex of the expression). Recently, texture descriptors and classification methods i.e. Local Binary Pattern (LBP) [17] and Local Phase Quantization (LPQ) [19] are also studied to extract appearance-based facial features. Lin et al. [33] proposed multistage discrimination model for facial expression recognition based on two-dimensional principal component analysis (2DPCA), and local texture represented by local binary pattern (LBP). They extensively tested their model on four databases and achieved promising results. Algorithm in [21] proposed extension of LBP descriptor for expression recognition. Both the methods achieved average expression recognition accuracy in the range of 96% for Cohn-Kanade facial expression database.

Geometric feature-based methods. For geometric feature-based methods [22–24], shapes and locations of facial components are extracted to form a feature vector. For expression recognition, Zhang et al. [22] has measured and tracked the facial motion using Kalman Filters. To achieve the recognition task they have also modeled the temporal behaviors of the facial expressions using Dynamic Bayesian networks (DBNs). In [23] authors have presented Facial Action Coding System’s (FACS) [34] Action Unit (AU) detection scheme by using features calculated from the particle filter tracked fiducial facial points. They trained the system on the MMI-Facial expression database [35] and tested on the Cohn-Kanade database [32] and achieved recognition rate of 84%. Bai et al. [24] extracted only shape information using Pyramid Histogram of Orientation Gradients (PHOG) and showed the "smile" detection accuracy as high as 96.7% on Cohn-Kanade database [32]. Research has been done with success in recent times to combine features extracted using appearance-based methods and geometric feature-based methods [36].

Recently researchers have shown great deal of interest in facial expression recognition from 3D (three dimension) face geometry. Most of the methods have been developed for facial expressions recognition from static 3D facial expression data [37–41]. However, recently community has proposed methods that employ dynamic 3D facial expression data for this purpose [42, 43]. Methods for 3D facial expression recognition usually consist of two main stages: feature extraction, and selection and classification of features, as in 2D face expression analysis. Dynamic systems for expression analysis also employ temporal modelling of the expression as a further step.

3 Visual experiment

The aim of visual experiment was to record the eye movement data of human observers in free viewing conditions. Data were analyzed in order to find which component of face is salient for specific displayed expression.

3.1 Methods

Eye movements of human observers were recorded as subjects watched a collection of 54 videos. Then saccades, blinks and fixations were segmented from each subject’s recording.

3.1.1 Participants and apparatus

Fifteen observers volunteered for experiment. They include both male and female aging from 20 to 45 years with normal or corrected to normal vision. All observers were naïve to the purpose of the experiment.

We used video based eye-tracker (Eyelink II system, SR Research) to record eye movements. The system consists of three miniature infrared cameras with one mounted on a headband for head motion compensation and the other two mounted on arms attached to headband for tracking both eyes. Each camera has a built-in infrared illuminator.

Stimuli were presented on a 19 inch CRT monitor with a resolution of 1024 x 768, and a refresh rate of 85Hz. A

Fig. 1 Six Universal expressions: first row show example images (Peak expression frame) from Cohn-Kanade (CK+) database [28], while second row show images from the Dartmouth database [30].

viewing distance of 70cm was maintained resulting in a 29° x 22° usable field of view as done by Jost et al. [44] and Khan et al. [45].

3.1.2 Stimuli

For the experiment, we used the videos from the extended Cohn-Kanade (CK+) database [28]. The CK+ database contains 593 sequences across 123 subjects which are FACS [34] coded at the peak frame. Out of 593 sequences only 327 sequences have emotion labels. This is because these are the only ones that fit the prototypic definition. Database consists of subjects aged from 18 to 50 years old, of which 69% were female, 81% Euro-American, 13% Afro-American and 6% others. Each video (without sound) showed a neutral face at the beginning and then gradually developed into one of six facial expressions. We selected 54 videos for the experiment. Videos were selected with the criteria that videos should show both male and female subjects, experiment session should complete within 20 minutes and posed facial expression should not look unnatural. Another consideration while selecting the videos was to avoid such sequences where the date/time stamp is not recorded over the chin of the subject. Figure 1 shows example of universal expressions with maximum intensity.

3.2 Procedure

The experiment was performed in a dark room with no visible object in observer's field of view except stimulus. It was carefully monitored that an experimental session should not exceed 20 minutes, including the calibration stage. This was taken care of in order to prevent participant's lose of interest or disengagement over time.

3.2.1 Eye movement recording

Eye position was tracked at 500 Hz with an average noise less than 0.01°. Fixations were estimated from a comparison between the center of the pupil and the reflection of the IR illuminator on the cornea. Each video was preceded by a black fixation cross displayed at the center of the screen on a uniform neutral gray background. This has a twofold impact: firstly all observers start viewing images from the same point and secondly, it allows gaze position to be realigned if headband slippage or significant pupil size change has deteriorated the accuracy of eye movement recording.

Head mounted eye-tracker allows flexibility to perform experiment in free viewing conditions as the system is designed to compensate for small head movements. Then the recorded data is not affected by head motions and participants can observe stimuli with no severe restrictions. Indeed, severe restrictions in head movements have been shown to alter eye movements and can lead to noisy data acquisition and corrupted results [46].

3.3 Visual Experiment: Results and Discussion

3.3.1 Gaze map construction

The most intuitively revealing output that can be obtained from the recorded fixations data is to obtain gaze maps. For every frame of the video and each subject i, the eye movement recordings yielded an eye trajectory T^i composed of the coordinates of the successive fixations f_k, expressed as image coordinates (x_k, y_k):

$$T^i = (f_1^i, f_2^i, f_3^i, ...) \tag{1}$$

As a representation of the set of all fixations f_k^i, a human gaze map $H(\mathbf{x})$ was computed, under the assumption that this map is an integral of weighted point spread functions

Fig. 2 Examples of gaze maps for six universal expressions. Each video sequence is divided in three mutually exclusive time periods. First, second and third columns show average gaze maps for the first, second and third time periods of a particular stimuli respectively.

$h(\mathbf{x})$ located at the positions of the successive fixations. It is assumed that each fixation gives rise to a normally (Gaussian) distributed activity. The width σ of the activity patch was chosen to approximate the size of the fovea. Formally, $H(\mathbf{x})$ is computed according to Equation 2:

$$H(\mathbf{x}) = H(x, y) = \sum_{i=1}^{N_{subj}} \sum_{f_k \in T^i} \exp\left(\frac{(x_k - x)^2 + (y_k - y)^2}{\sigma^2}\right) \quad (2)$$

where (x_k, y_k) are the spatial coordinates of fixation f_k, in image coordinates. In Figure 2 gaze maps are presented as the heat maps where the colored blobs / human fixations are superimposed on the frame of a video to show the areas where observers gazed. The longer the gazing time is, the warmer the color is.

As the stimuli used for the experiment is dynamic i.e. video sequences, it would have been incorrect to average all the fixations recorded during trial time (run length of video) to construct gaze maps as this could lead to biased analysis of the data. To meaningfully observe and analyze the gaze trend across one video sequence, we have manually divided each video sequence in three mutually exclusive time periods. The first time period correspond to initial frames of the video sequence where the actor's face has no expression i.e. neutral face. The last time period encapsulates the frames where the actor is showing expression with full intensity (apex frames). The second time period is a encapsulation of frames which has a transition of facial expression i.e. transition from the neutral face to the beginning of the desired expression (i.e neutral to onset of the expression). Then the fixations recorded for a particular time period are averaged across 15 observers.

Gaze maps presented in the Figure 2 suggests the saliency of mouth region for the expressions of happiness and surprise. It can be observed from the figure that as the two said expressions becomes prominent(second and third time periods) most of the gazes are attracted towards only one facial region and that is the mouth region. The same observation can be made for the facial expressions of sadness and fear but with some doubts. For the expressions of anger and disgust it seems from the gaze maps that no single facial region emerged as salient, as the gazes are attracted towards two to three facial regions even when the expression was show at its peak.

3.3.2 Substantiating observations through statistical analysis

In order to statistically confirm the intuition gained from the gaze maps about the saliency of different facial region(s) for the six universal facial expressions we have calculated the average percentage of trial time observers have fixated their gazes at specific region(s) in a particular time period (definition of time period is same as described previously). The resulting data is plotted in Figure 3.

Figure 3 shows that the region of mouth is the salient region for the facial expressions of happiness and surprise. This result is consistent with the results shown by Cunningham et al. [47], and Boucher et al. [48]. It can be easily observed from the figure that, as the expressions become more prominent (third period), the humans tend to fixate their gazes mostly at the region of mouth. This observation also holds for the expression of sadness.

Facial expression of disgust shows random behavior. Even when stimuli show expression with maximum

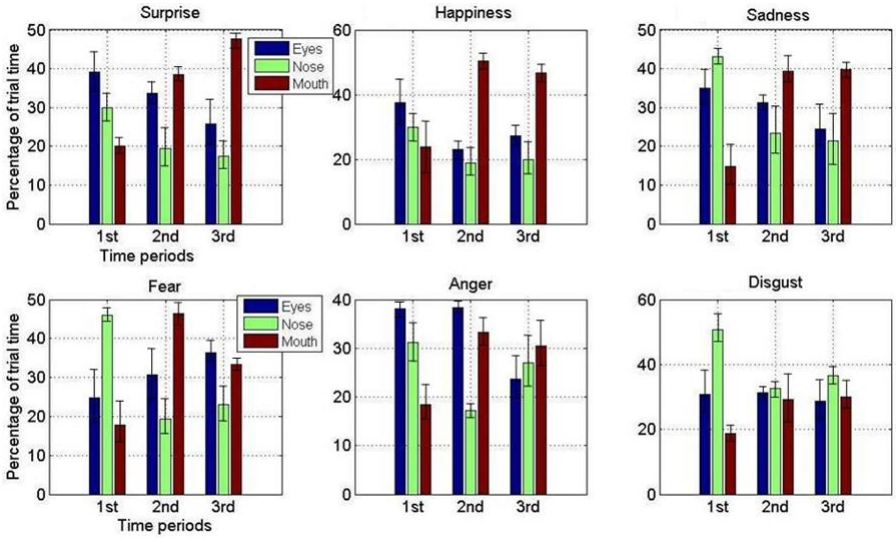

Fig. 3 Average percentage of trial time observers have spent on gazing different facial regions in a particular time period. The error bars represent the standard error (SE) of the mean. First time period: initial frames of video sequence. Second time period: frames which has a transition from neutral face to particular expression. Third time period: apex frames.

intensity, observers have gazed all the three regions randomly (see Figure 3, third time period).

In expression of fear facial regions of mouth and eye attract most of the gazes. From Figure 3 it can be seen that in second time period (period correspond to the time when observe experiences the change in face presented in stimuli toward the maximum expression) observers mostly gazed at the mouth region and in the final trial period eye and mouth regions attract most of the attention.

In 1975 Boucher et al. [48] wrote that "anger differs from the other five facial expressions of emotion in being ambiguous" and this observation holds for the current study as well. Anger shows complex interaction of eyes, mouth and nose regions without any specific trend. Conclusions drawn from the experimental study are summarized in Table 1.

Table 1 Summary of the facial regions that emerged as salient

Facial expression	Salient facial region(s)
Happiness	Mouth region.
Surprise	Mouth region.
Sadness	Mouth and eye regions. Biased towards mouth region.
Disgust	Nose, mouth and eye regions. Wrinkles on the nose region gets little more attention than the other two regions.
Fear	Mouth and eye regions.
Anger	Mouth, eye and nose regions.

4 Novel framework for automatic facial expressions recognition

Results of the visual experiment provided the evidence that human visual system gives importance to three regions i.e. eyes, nose and mouth, while decoding six universal facial expressions. In the same manner, we argue that the task of expression analysis and recognition could be done in more conducive manner, if same regions are selected for further processing. We propose to extract three features only from the salient regions of face. These three features are entropy, brightness and multi-resolution Local Binary Pattern (LBP) [27].

Entropy is the measure of uncertainty or measure of absence of the information associated with the data [49] while brightness as described by Wyszecki and Stiles [50] is an *"attribute of a visual sensation according to which a given visual stimulus appears to be more or less intense"*. Currently, there is no standard or reference formula for brightness calculation. We propose to use BCH (Brightness, Chroma, Hue) model [51] for brightness calculation. LBP features were initially proposed for texture analysis [17], but recently they have been successfully used for facial expression analysis [21, 52]. Overview of the proposed framework is summarized in algorithm 1 and the details related to each step is presented in a subsequent subsections.

The rationale behind extracting entropy and brightness features is that, as a first step framework calculates saliency map which show salient regions by highlighting them. By extracting brightness feature, framework calculates local saliency value (brightness is proportional to saliency).

Algorithm 1: Proposed framework for facial expression recognition

input : video frame
output: expression label

for $i \leftarrow 1$ **to** $numFrames$ **do**
- calculate saliency map using "frequency-tuned salient region detection" algorithm [12]
- automatically localize salient facial regions using viola-jones algorithm [13]
- calculate entropy [49] from salient region using: $E = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$
- calculate brightness features from salient region using BCH model [51]: $B = \sqrt{D^2 + E^2 + F^2}$
- Calculate multi-resolution Local Binary Pattern (LBP) features [27] from salient regions
- concatenate entropy,brightness and LBP feature to make feature vector, i.e. $f = \{ E, B, LBP \}$
- classify input stimuli to one of six universal expression using feature vector f

Entropy values explains how well particular facial region is mapped as salient. While LBP features provide global facial texture information which is very significant for facial expression analysis.

4.1 Salient region detection

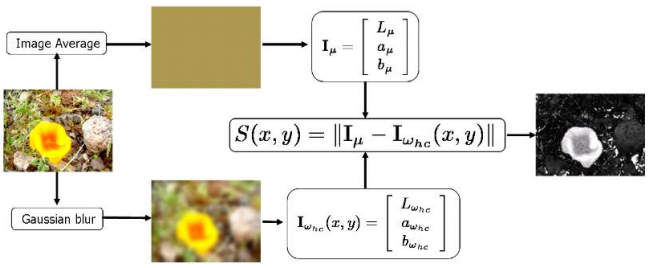

Fig. 4 Overview of frequency-tuned salient region detection algorithm. [12]

First problem that is addressed, is to find salient region detection algorithm that produces saliency maps which detects similar salient facial regions as concluded by the psycho-visual experiment. The idea is to use saliency map information to find appropriate weights (or value for brightness and entropy) for extracted features. In order to find appropriate saliency detection algorithm we examined four state-of-the-art methods [12, 53–55].

Four state-of-the-art methods for extracting salient regions are "Itti's model" by Itti et al. [53], "frequency tuned salient region detection" by Achanta et al. [12], "graph-based visual saliency" by Harel et al. [54] and "spectral residual approach for saliency detection" by Hou and Zhang [55] referred here as IT, FT, GBVS and SR respectively. The choice of these algorithms is motivated by the following reasons:

1. Citation in literature. The classic approach of IT is widely cited.
2. Recency. GBVS, SR and FT are recent.
3. Variety. IT is biologically motivated,GBVS is a hybrid approach, FT and SR estimates saliency in the frequency domain, and FT outputs full-resolution saliency maps.

We propose to use *frequency-tuned salient region detection* algorithm (referred as FT later in the text) developed by Achanta et al. [12] for detection and localization of salient facial regions. We have chosen this model over other existing state-of-the-art models because it performs better in predicting human fixations (see Figure 6 and Figure 5).

Figure 5 shows that except FT model, none of the other examined model correctly predicts human gazes. Itti method and GBVS method outputs quite similar saliency maps with wrong predictions, while SR method outputs saliency map that is only 64 x 64 pixels in size with no significant correct prediction.

Figure 6 shows salient regions for six expressions as detected by FT. It can be observed from the figure that most of the time it predicts three regions as salient facial region i.e. nose, mouth and eyes which is in accordance with visual experiment result. Secondly, a distinctive trend in detected salient regions and associated brightness can also be observed for different expressions. Another advantage of the FT model is its computational efficiency, which is very important for the system to run in real time. Lastly, this model outputs saliency maps in full resolution, which is not the case for SR model. Due to all of these benefits we concluded to use frequency-tuned salient region detection algorithm developed by Achanta et al. [12] for detection of salient facial regions in our framework.

FT algorithm finds low-level, pre-attentive, bottom-up saliency. Biological concept of center-surround contrast is the core of this algorithm, but is not based on any biological model. Frequency-tuned approach is used to estimate center-surround contrast using color and luminance features. According to Achanta et al. [12] it offers three advantages over existing methods: uniformly highlighted salient regions with well defined boundaries, full resolution saliency maps, and computational efficiency. This algorithm find the Euclidean distance between pixel vector in a Gaussian filtered image (2D Gaussian filter is given by Equation 3) with the average vector in a Lab color space [50]. This is illustrated in the Figure 4.

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (3)$$

Where σ (σ=3 in method proposed by Achanta et al. [12]) is the standard deviation of the distribution.

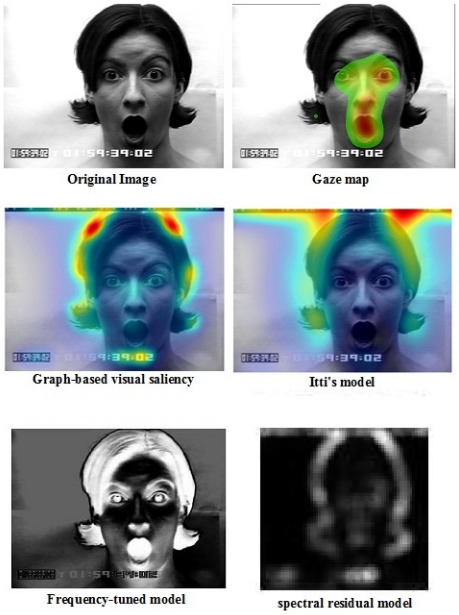

Fig. 5 Comparison of automatic detected salient regions with gaze map obtained from psycho-Visual experiment (see Section 3 for reference). First row: original image and average gaze map of fifteen observers. Second row: saliency maps obtained from GBVS [54] and Itti methods [53] (shown as heat maps). Third row: saliency maps obtained from FT [12] and SR methods [55].

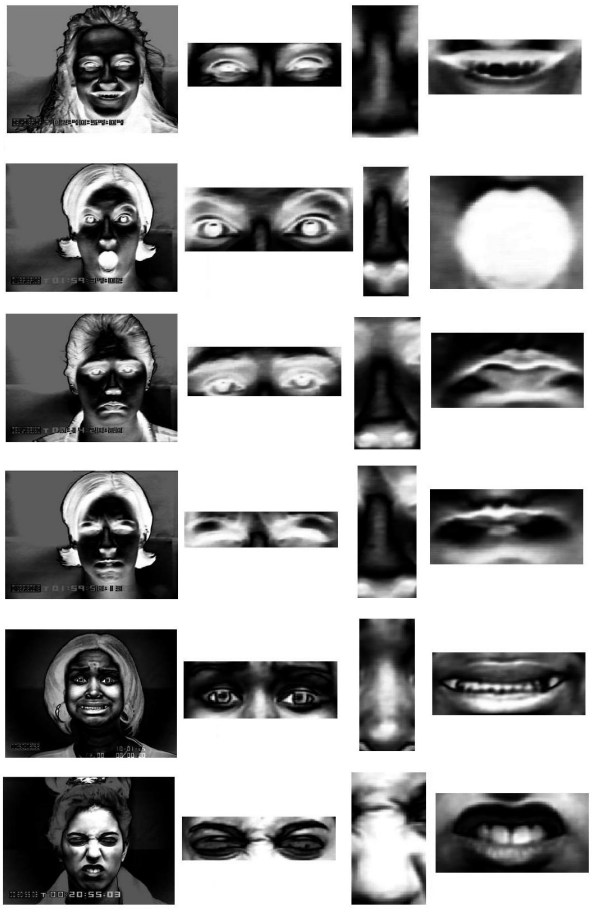

Fig. 6 Salient region detection for different expressions using FT [12]. Each row shows detected salient regions in a complete frame along with the zoom of three facial regions i.e. eyes, nose and mouth. Brightness of the salient regions is proportional to its saliency. First row shows expression of happiness, second row: surprise, third row: sadness, fourth row: anger, fifth row: fear and sixth row: disgust .

4.2 Feature Extraction

We have chosen to extract the features of LBP, brightness and entropy for automatic recognition of expressions as these features have shown a discriminative trend. For entropy and brightness features discriminative trend can be observed from Figure 6 and 8. Figure 8 shows the average entropy values for the facial regions. Each video is manually divided in three equal time periods for the reasons discussed earlier (see Section 3). The entropy values for the facial regions corresponding to specific time periods (definition of time periods is same as discussed earlier) are averaged and plotted in Figure 8.

Apart from using these two features to automatically recognize facial expressions, we are also proposing to extract appearance features to build robust feature vector. We propose to extract local binary pattern (LBP) [17] features from salient regions of face and concatenate them with brightness and entropy features.

By using FT saliency model, we obtained saliency maps for every frame of the video. Then obtained saliency maps are further processed for the calculation of brightness and entropy value for the three facial regions.

4.2.1 Brightness Calculation

Brightness is one of the most significant pixel characteristics but currently, there is no standard formula for brightness calculation. The brightness values, as explained earlier, are calculated using BCH (Brightness, Chroma, Hue) model [51]. For B (Brightness) C (Chroma) H (Hue) color coordinate system (CCS), the following definitions for Brightness, Chroma and Hue are used:

1. B: a norm of a color vector S.
2. C: an angle between the color vector S and an axis D - color vector representing Day Light (for example D65, D55, EE etc.).
3. H: is the angle between the orthogonal projection of the color vector S on the plane orthogonal to the axis D and an axis E - the orthogonal projection of a color vector, corresponding to some fixed stimulus (for example, a monochromatic light with wavelength 700 nm), on the same plane.

Figure 7 illustrates the relationship between values D, E, and F (E = F = 0 for grey color in DEF color coordinate system), coordinates of the vector S in an orthogonal coordinate system DEF, and parameters B, C, and H, which

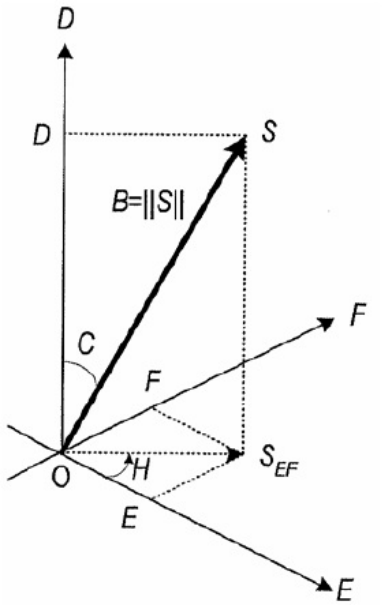

Fig. 7 Color Coordinate Systems DEF and BCH [51]

might be considered as spherical coordinates of the vector S in BCH coordinate system.

Use of stimulus length as a measure of Brightness introduced in BCH (Brightness, Chroma, Hue) model provides Brightness definition effective for different situations. Length is calculated according to Cohen metrics [56].

$$B = \sqrt{D^2 + E^2 + F^2} \tag{4}$$

$$\begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 0.2053 & 0.7125 & 0.4670 \\ 1.8537 & -1.2797 & -0.4429 \\ -0.3655 & 1.0120 & -0.6104 \end{bmatrix} . \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{5}$$

where X, Y, and Z are tristimulus values [57]. Following equation is used to convert RGB value to XYZ color space value:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5943 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{6}$$

The main advantage of BCH model is that it performs only intended operation without unwilling concurrent modification/processing of other image parameters.

4.2.2 Entropy Calculation

The entropy values for eyes, nose and mouth regions are calculated using equation 7:

$$E = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \tag{7}$$

where n is the total number of grey levels, and $p = \{p(x_1), ..., p(x_n)\}$ is the probability of occurrence of each level.

In the context of this article, we have used entropy as a measure to quantitatively determine whether a particular facial region is fully mapped as salient or not. Higher value of entropy for a particular facial region corresponds to higher uncertainty or points out the fact that the facial region is not fully mapped as salient. To calculate entropy value stimuli is first converted in grayscale image.

From Figure 8 it can be observed that the average entropy values for the region of mouth, for the expressions of happiness and surprise are very low as compared to entropy values for the other regions. This finding shows that the region of mouth was fully mapped (can also be seen in Figure 6) as salient by saliency model, and the same we concluded from our visual experiment. It is also observable from the figure that the values of entropy for the region of mouth for these two expressions is lower than any other entropy values for the rest of facial expressions in the second and third time periods. This result shows that there is a discriminative trend in entropy values which will help in automatic recognition of facial expressions.

Entropy values for the expression of sadness show discriminative trend and suggests that nose and mouth regions are salient with more biasness towards mouth region. This results conforms very well with the results from visual experiment.

For the expression of disgust, entropy value for the facial region of nose is quite low pointing out the fact that the region of nose is mapped fully as salient which again is in accordance with our visual experiment result. This conclusion can also be exploited for the automatic facial expression recognition of disgust.

We obtained low entropy value for the facial region of eyes for the expression of anger. This points to the fact that according to the saliency model the region of eyes emerges as salient. But the results from the visual experiment show complex interaction of all three regions. Entropy values for the expression of fear also show different result from the visual experiment. The discrepancies found in the entropy values for the expressions of anger and fear are neither negligible nor significant and will be studied and addressed in future work.

4.2.3 Local Binary Pattern

Local Binary Pattern (LBP) features were initially proposed for texture analysis [17], but recently they have been successfully used for facial expression analysis [21, 52, 58]. The most important property of LBP features is their tolerance against illumination changes and their computational simplicity [17, 59]. The operator labels the pixels of an image by thresholding the 3 x 3 neighbourhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels can be

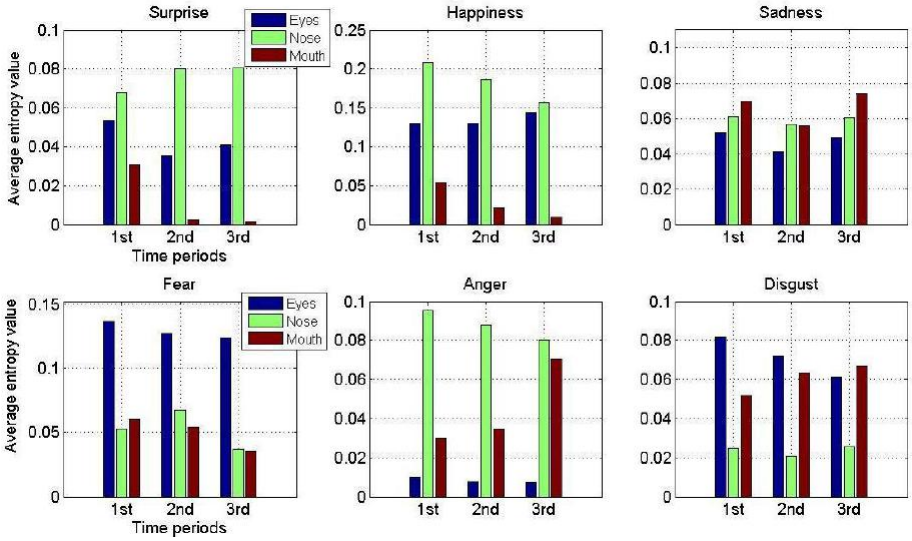

Fig. 8 Average entropy value for different facial regions. First time period: initial frames of video sequence. Third time period: apex frames. Second time period: frames which has a transition from neutral face to particular expression.

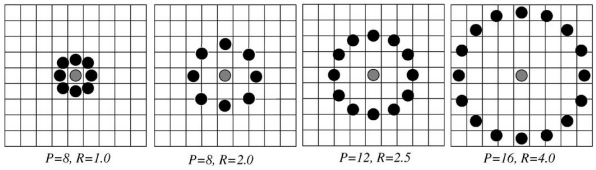

Fig. 9 Examples of the extended LBP [59]. The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel

used as a texture descriptor. Formally, LBP operator takes the form:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \tag{8}$$

where in this case n runs over the 8 neighbours of the central pixel c, i_c and i_n are the grey level values at c and n and $s(u)$ is 1 if $u \geqslant 0$ or 0 otherwise.

The limitation of the basic LBP operator is its small 3 x 3 neighborhood which can not capture dominant features with large scale structures. Hence the operator later was extended to use neighborhood of different sizes [59]. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. See Figure 9 for examples of the extended LBP operator, where the notation (P, R) denotes a neighborhood of P equally spaced sampling points on a circle of radius of R that form a circularly symmetric neighbor set.

The LBP operator with P sampling points on a circular neighborhood of radius R is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{9}$$

where, g_c is the gray value of the central pixel, g_p is the value of its neighbors, P is the total number of involved neighbors and R is the radius of the neighborhood.

Another extension to the original operator is the definition of uniform patterns, which can be used to reduce the length of the feature vector and implement a simple rotation-invariant descriptor. This extension was inspired by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. For example, 00000000, 00011110 and 10000011 are uniform patterns. Accumulating the patterns which have more than 2 transitions into a single bin yields an LBP operator, denoted $LBP^{u2}_{P,R}$ patterns. These binary patterns can be used to represent texture primitives such as spot, flat area, edge and corner.

In our experiment (see next Section) we extracted multi-resolution LBP features by varying radius R [27]. We extracted multi-resolution LBP features from salient region using two scales, which are $LBP^{u2}_{8,1}$ and $LBP^{u2}_{16,2}$ (See Figure 9 for reference). $LBP^{u2}_{8,1}$ denotes a uniform LBP operator with 8 sampling pixels in a local neighborhood region of radius 1. While $LBP^{u2}_{16,2}$ denotes a uniform LBP operator with 16 sampling pixels in a local neighborhood region of radius 2. By employing multi-resolution LBP operator framework extracts features from fine as well as crude stimuli level, which is important to robustly recognize facial expressions.

5 Automatic Expression Recognition Experiments

To measure the performance of the proposed framework for facial expression recognition we conducted experiments using three databases:

1. Extended Cohn-Kanade (CK+) database [28] (posed expressions)
2. FG-NET FEED [29] database (natural expressions)
3. Dartmouth database of children's faces [30]

For all the experiments, the performance of the framework was evaluated using classical classifier i.e. "Support vector machine (SVM)" with χ^2 kernel and γ=1. SVM performs an implicit mapping of data into a higher dimensional feature space, and then finds a linear separating hyperplane with the maximal margin to separate data in this higher dimensional space. Given a training set of labeled examples $\{ (x_i, y_i) , i = 1 \ldots l \}$ where $x_i \in \Re^n$ and $y_i \in \{-1, 1\}$, a new test example x is classified by the following function:

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b) \quad (10)$$

where α_i are Langrange multipliers of a dual optimization problem that describe the separating hyperplane, $K(.,.)$ is a kernel function, and b is the threshold parameter of the hyperplane. We used Chi-Square kernel as it is best suited for histograms. It is given by:

$$K(x, y) = 1 - \sum_i \frac{2 \times (x_i - y_i)^2}{(x_i + y_i)} \quad (11)$$

Average recognition accuracy / rate is calculated using *10*-fold cross validation technique. In k-fold cross validation, features vector set is divided into k equal subsets. k-1 subsets are used for the training while a single set is retained for the testing. The process is repeated k times (k-folds), with each of the k subsets used exactly once for testing. Then, the k estimations from k-folds are averaged to produce final estimated value.

5.1 Experiment on the extended Cohn-Kanade (CK+) database

For this experiment we used all 309 sequences from the CK+ database (See Section 2 for reference) which have FACS coded expression label [34]. The experiment was carried out on the frames which covers the status of onset to apex of the expression, as done by Yang et al. [16]. Region of interest (salient facial regions) was obtained automatically by using Viola-Jones object detection algorithm [13] and processed to obtain feature vector. As per Section 3, we concluded that HVS is mostly attracted toward three facial region. Thus frameworks extracts all these features from three facial regions i.e. eyes, nose and mouth and concatenates them to build final feature vector. The proposed framework achieved average recognition rate of 94.9% for six universal facial expressions using *10*-fold cross validation.

Table 2 Comparison with the state-of-the-art methods for Cohn-Kanade database.

	Sequence Num	Class Num	Performance Measure	Recog. Rate (%)
[14]	313	7	leave-one-out	93.3
[21]	374	6	2-fold	95.19
[21]	374	6	10-fold	96.26
[36]	374	6	5-fold	94.5
[58]	309	6	10-fold	96.7
[15]	375	6	-	93.8
[16]a	352	6	66% split	92.3
[16]b	352	6	66% split	80
Ours	**309**	**6**	**10-fold**	**94.9**

Table 2 shows the comparison of the achieved average recognition rate of the proposed framework with the state-of-the-art methods using same database (i.e Cohn-Kanade database). It is evident from the Table that proposed framework achieved results at par with state-of-the-art methods. Results from [16] are presented for the two configurations. " [16]a" shows the result when the method was evaluated for the last three frames from the sequence while " [16]b" presents the reported result for the frames which encompasses the status from onset to apex of the expression. The method discussed in " [16]b" is directly comparable to our method (frames which covers the status of onset to apex of the expression). It can be observed from the Table 2 that the proposed framework is comparable to any other state-of-the-art method in terms of expression recognition accuracy. The method discussed in " [16]b" is directly comparable to our method (frames which covers the status of onset to apex of the expression). In this configuration, our framework achieved better recognition accuracy with relatively very small feature vector.

5.2 Experiment on the FG-NET FEED database

FG-NET FEED contains 399 video sequences across 18 different individuals showing seven facial expressions i.e. six universal expression [9] plus one neutral. In this database individuals were not asked to act rather expressions were captured while showing them video clips or still images to wake real expressions.

The proposed framework achieved average recognition rate of 86.7% for six universal facial expressions using *10*-fold cross validation. Table 3 shows the comparison of the achieved average recognition rate of the proposed framework with the state-of-the-art methods using same database (i.e FG-NET FEED database).

Table 3 Comparison with the state-of-the-art methods for FG-NET FEED data

	Class Num	Performance Measure	Recog. Rate (%)
[58]	6	10-fold	92.3
[61]	7	-	82
[62]	7	10-fold	70
[63]	7	ROC	90.3
[64]	5	-	81.7
Ours	**6**	**10-fold**	**86.7**

5.3 Experiment on stimuli containing children's faces

To test the effectiveness of the proposed framework on children faces, we conducted experiment on the the Dartmouth database [30]. The Dartmouth Database of children Faces contains faces of 40 male and 40 female Caucasian children (see Figure 1 show example images from the database). All faces in the database were assessed by at least 20 raters for facial expression identifiability and intensity (as opposed to CK+ database which is FACS [34] compliant/coded). Expression of happy was most accurately identified while fear was least accurately identified by human raters. Human raters correctly classified 94.3% of the happy faces while expression of fear was correctly identified in 49.08% of the images, least identifiable by human raters. On average human raters correctly identified expression in 79.7% of the images [30].

For the experiment we used all the frames in the database. Region of interest was obtained automatically by using Viola-Jones object detection algorithm [28] and processed to extract proposed features. Proposed framework achieved average recognition rate of 82.3% using *10*-fold cross validation.

Proposed framework achieved recognition accuracy of more than 94% for CK+ database (refer Section 1) but for Dartmouth database of children's faces it achieved average recognition accuracy of 82.3% . This is due to the fact that database of children faces [30] have actors ranging from 6 to 16 years of ages. Usually children show expressions in a subtle way, thus creating large inter and intra population variations as opposed to adults. This fact is visible in Figure 10. It is observable in the figure that same expression is produced by different kids in completely different manner and thus making very difficult for classifier to learn them robustly.

5.3.1 Generalization Capabilities

Aim of this experiment is to study how well the proposed framework generalizes on unseen data or new database. According to our knowledge only Valstar et al. [23] have reported such data earlier. In this experiment we trained classifier using the Dartmouth database [30] and tested its performance on frames from NIMH child emotional faces picture set (NIMH-ChEFS) database [60]. NIMH-ChEFS database has 482 frames containing expressions of fear, anger, happy and sad with two gaze conditions: direct and averted gaze. The databases is validated by 20 adult raters.

Fig. 10 Example of inconsistent expression shown by actors in the Dartmouth database of children faces [30]. First row show images that have a ground truth label of anger. Second row show images that have label of disgust, while third row present images that have label of fear.

Fig. 11 Expressions as shown in NIMH child emotional faces picture set (NIMH-ChEFS) database [60]. First four images show example with averted gaze while second four images with direct gaze.

Proposed framework achieved average recognition accuracy of 82.3% when trained on the Dartmouth database [30]. While it achieved average recognition accuracy of 76.8% when it was tested on NIMH-ChEFS database [60]. It is important to note that training and testing samples were completely different as they came from two different databases. This experiment simulates the real life situation when the proposed framework would be employed to recognize facial expressions on the unseen data. Obtained results are encouraging and they can be further improved by training classifiers on more than one databases before using in real life scenario.

6 Conclusion

The experimental study presented in this article provides the insight into which facial region(s) emerges as salient according to human visual attention for six universal expressions. Eye movements of fifteen observers were recorded using an eye-tracker as they watched the stimuli showing facial expressions. The analysis of data revealed the fact that for six universal facial expressions, visual attention

is mostly grabbed by three facial regions i.e. eyes, mouth and nose regions. For the expressions of happiness and surprise, the facial region of mouth emerged as salient. Expression of sadness shows the same result with little more attention towards the region of eyes. The regions of eyes and mouth captures most of the gazes for the expressions of fear while the expressions of anger and disgust show the complex interaction of mouth, nose and eyes regions.

Secondly, we show that facial expressions can be recognized automatically by imitating human visual system. Proposed framework utilizes very well know saliency detection model along with the measure of LBP, entropy and brightness. According to our knowledge no scientist has exploited the measure of brightness and entropy to recognize facial expressions. In the future, we will extend the proposed framework so that it can recognize a wide array of expressions. We will focus on incorporating movement information in our descriptor to make it more accurate and robust. Research is required to be done to recognize expressions across camera angle variations.

References

1. P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Pergamon Press, New York, 1972.
2. P. Ekman and W. V. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Prentice Hall, Englewood Cliffs, New Jersey, 1975.
3. P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W. W. Norton & Company, New York, 3rd edition, 2001.
4. P. Ekman. Facial expression of emotion. *Psychologist*, 48:384–392, 1993.
5. P. Bull. State of the art: Nonverbal communication. *Psychologist*, 14:644–647, 2001.
6. V. H. Yngve. On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–578. Chicago Linguistic Society, 1970.
7. P. Carrera-Levillain and J. Fernandez-Dols. Neutral faces in context: Their emotional meaning and their function. *Journal of Nonverbal Behavior*, 18:281–299, 1994.
8. J. FernÃąndez-Dols, H. Wallbott, and F. Sanchez. Emotion category accessibility and the decoding of emotion from facial expression and context. *Journal of Nonverbal Behavior*, 15:107–123, 1991.
9. P. Ekman. Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, pages 207–283. Lincoln University of Nebraska Press, 1971.
10. U. Rajashekar, L. K. Cormack, and A.C Bovik. Visual search: Structure from noise. *Eye Tracking Research & Applications Symposium*, pages 119–123, 2002.
11. R.A. Khan, H. Konik, and E. Dinet. Enhanced image saliency model based on blur identification. *IEEE International Conference of Image and Vision Computing New Zealand (IVCNZ)*, pages 1 –7, 2010.
12. Achanta R., Hemami S., Estrada F., and Susstrunk S. Frequency-tuned salient region detection. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2009.
13. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
14. G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24:615–625, 2006.
15. Y. Tian. Evaluation of face resolution for expression analysis. *Computer Vision and Pattern Recognition Workshop*, 2004.
16. P. Yang, Q. Liu, and D. N. Metaxas. Exploring facial expressions with compositional features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
17. T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 29:51–59, 1996.
18. R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz. Exploring human visual system: study to aid the development of automatic facial expression recognition framework. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012.
19. V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. *International conference on Image and Signal Processing*, 2008.
20. R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz. Human vision inspired framework for facial expressions recognition. *IEEE International Conference on Image Processing (ICIP)*, 2012.
21. G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 29:915–928, 2007.
22. Y. Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:699–714, 2005.
23. M.F. Valstar, I. Patras, and M. Pantic. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 76–84, 2005.
24. Y. Bai, L. Guo, L. Jin, and Q. Huang. A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition. *IEEE International Conference on Image Processing (ICIP)*, 2009.
25. R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz. Pain detection through shape and appearance features. *IEEE International Conference on Multimedia and Expo (ICME)*, 2013.
26. R. A. Khan, A. Meyer and S. Bouakaz. Automatic affect analysis: from children to adults. *International Symposium on Visual Computing (ISVC)*, 2015.
27. G. Zhao and M. Pietikäinen. Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. *Pattern Recognition Letters*, 30 : 1117–1127, 2009.
28. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kande dataset (CK+): A complete fa-

cial expression dataset for action unit and emotion-specified expression. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2010.

29. F. Wallhoff. Facial expressions and emotion database. www.mmk.ei.tum.de/~waf/fgnet/feedtum.html, 2006.
30. Dalrymple, K.A., Gomez, J., Duchaine, B. *The Dartmouth database of children's faces: Acquisition and validation of a new face stimulus set. PLoS ONE*, 8, 2013.
31. Y. Tian, T. Kanade, and J. F. Cohn. *Handbook of Face Recognition*. Springer, 2005 (Chapter 11. Facial Expression Analysis).
32. T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. *IEEE International Conference on Automatic face and Gesture Recognition (FG'00)*, pages 46–53, 2000.
33. Daw-Tung Lin and De-Cheng Pan. Integrating a mixed-feature model and multiclass support vector machine for facial expression recognition. *Integr. Comput.-Aided Eng.*, 16(1):61–74, January 2009.
34. P. Ekman and W. Friesen. The facial action coding system: A technique for the measurement of facial movements. *Consulting Psychologist*, 1978.
35. M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. *IEEE International Conference on Multimedia and Expo*, 2005.
36. I. Kotsia, S. Zafeiriou, and I. Pitas. Texture and shape information fusion for facial expression and facial action unit recognition. *Pattern Recognition*, 41:833–851, 2008.
37. X. Yang and D. Huang and Y. Wang and L. Chen. Automatic 3D facial expression recognition using geometric scattering representation. *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 2015.
38. X. Zhao and D. Huang and E. Dellandrea and L. Chen. Automatic 3D Facial Expression Recognition Based on a Bayesian Belief Net and a Statistical Facial Feature Model. *International Conference on Pattern Recognition*, 2010.
39. Arman Savran and Bulent Sankur and M. Taha Bilge. Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units. *Pattern Recognition*, 45 : 767–782, 2012.
40. Lijun Yin and Xiaozhou Wei and P. Longo and A. Bhuvanesh. Analyzing Facial Expressions Using Intensity-Variant 3D Data For Human Computer Interaction. *18th International Conference on Pattern Recognition*, 2006.
41. H. Li and H. Ding and D. Huang and Y. Wang and X. Zhao and J. Morvan and L. Chen. An efficient multimodal 2D + 3D feature-based approach to automatic facial expression recognition. *Computer Vision and Image Understanding*, 140 : 83–92, 2015.
42. M. Rosato and X. Chen and L. Yin. Automatic Registration of Vertex Correspondences for 3D Facial Expression Analysis. *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2008.
43. V. Le and H. Tang and T. S. Huang. Expression recognition from 3D dynamic faces using robust spatio-temporal shape features. *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 2011.
44. T. Jost, N. Ouerhani, R. Wartburg, R. Müri, and H. Hügli. Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding. Special Issue on Attention and Performance in Computer Vision*, 100:107–123, 2005.
45. R.A. Khan, H. Konik, and E. Dinet. Visual attention: effects of blur. *IEEE International Conference on Image Processing (ICIP), Brussels Belgium*, 2011.
46. H. Collewijin, M. R. Steinman, J. C. Erkelens, Z. Pizlo, and J. Steen. *The Head-Neck Sensory Motor System*. Oxford University Press, 1992.
47. D. W. Cunningham, M. Kleiner, C. Wallraven, and H. H. Bülthoff. Manipulating video sequences to determine the components of conversational facial expressions. *ACM Transactions on Applied Perception*, 2:251–269, 2005.
48. J. D. Boucher and P. Ekman. Facial areas and emotional information. *Journal of communication*, 25:21–29, 1975.
49. C. E. Shannon and W. Weave. *The Mathematical Theory of Communication*. University of Illinois Press, 1963.
50. G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. JohnWiley and Sons, New York., 1982.
51. S. Bezryadin and P. Bourov. Color coordinate system for accurate color image editing software. *International Conference Printing Technology*, pages 145–148, 2006.
52. C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27:803–816, 2009.
53. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20, pages 1254–1259, 1998.
54. C. Koch J. Harel and P. Perona. Graph-based visual saliency. *Proceedings of Neural Information Processing Systems (NIPS)*, 2006.
55. X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
56. Cohen J. B. *Visual Color and Color Mixture: The Fundamental Color Space*. University of Illinois Press, 2000.
57. *Commission internationale de l'Eclairage proceedings (CIE)*. Cambridge University Press, Cambridge., 1931.
58. R.A. Khan, A. Meyer, H. Konik, and S. Bouakaz. Framework for reliable, real-time facial expression recognition for low resolution images . *Pattern Recognition Letters*, 34(10):1159–1168, 2013.
59. T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24:971–987, 2002.
60. Egger, H., Pine, D., Nelson, E., Leibenluft, E., Ernst, M., K.E., T., Angold, A. The NIMH child emotional faces picture set (NIMH-ChEFS): A new set of children's facial emotion stimuli. *Int. J. Methods Psychiatr. Res.*, 20:145–56, 2011.
61. C. Zhan, W. Li, P. Ogunbona, and F. Safaei. A real-time facial expression recognition system for online games. *International Journal of Computer Games Technology*, 2008.
62. Ewa Piatkowska. Facial expression recognition system. Master's thesis, DePaul University, College of Computing and Digital Media, Chicago USA, 2010.

63. A. Khanum, M. Mufti, Y. Javed, and Z. Shafiq. Fuzzy case-based reasoning for facial expression recognition. *Fuzzy Sets and Systems*, 160(2):231 – 250, 2009.
64. S. Rosdiyana and S. Hideyuki. Extraction of the minimum number of gabor wavelet parameters for the recognition of natural facial expressions. *Artificial Life and Robotics*, 16(1):21–31, 2011.

Rizwan Ahmed Khan is an Associate Professor at Barrett Hodgson University, Pakistan. He has received PhD in Computer Science from Université Claude Bernard Lyon 1, France in 2013. He has worked as postdoctoral research associate at Laboratoire d'InfoRmatique en Image et Systemes d'information (LIRIS), Lyon, France. His research interests include computer vision, image processing, pattern recognition and human perception.

Alexandre Meyer received his PhD degree in Computer Science from Université Grenoble 1, France in 2001. From 2002 to 2003, he was postdoctoral fellow at University College London. Since 2004 he is working as Associate Professor at Université Claude Bernard Lyon 1, France and member of the LIRIS research lab. His current research concerns Computer Animation and Computer Vision of characters.

Hubert Konik received his PhD degree in Computer Science from Université Jean Monnet in 1995. He is Associate Professor at Télécom Saint-Etienne and member of Image Science & Computer Vision team of Laboratoire Hubert Curien, Saint-Etienne, France. His research interests are focused on image processing and analysis, more particularly content aware image processing for new services and usages.

Saida Bouakaz received her PhD from Joseph Fourier University in Grenoble, France. She is Full Professor at the department of Computer Science, Université Claude Bernard Lyon 1, France. Her research interests include computer vision and graphics including motion capture and analysis, gesture recognition and facial animation.