Une méthode de reconnaissance des expressions du visage basée sur la perception
Human perception based framework for automatic facial expression recognition

Rizwan Ahmed Khan[1,2], Alexandre Meyer[1,2], Hubert Konik[1,3], Saida Bouakaz[1,2]

[1] Université de Lyon, CNRS
[2] Université Lyon 1, LIRIS, UMR5205, F-69622, France
[3] Université Jean Monnet, Laboratoire Hubert Curien, UMR5516, 42000 Saint-Etienne, France

{*Rizwan-Ahmed.Khan, Alexandre.Meyer, Saida.Bouakaz*}*@liris.cnrs.fr, hubert.konik@univ-st-etienne.fr*

Résumé

Les humains peuvent reconnaître très facilement les expressions du visage en temps réel. Toutefois, la reconnaissance fiable et rapide des expressions faciales en temps réel est une tâche difficile pour un ordinateur. Nous présentons une nouvelle approche de reconnaissance de trois type d'expressions faciales qui se base sur l'idée de ne considérer que de petites régions du visage bien définies pour en extraire les caractéristiques. Cette proposition est basée sur une étude psycho-visuel expérimental menée avec un eye-tracker. Les mouvements des yeux de quinze sujets ont été enregistrés dans des conditions de visualisation libre d'une collection de 54 vidéos montrant six expressions faciales universelles. Les résultats de cette étude montrent que pour certaines expressions du visage une unique région est perceptuellement plus attractive que les autres. Les autres expressions montrent une attractivité pour deux ou trois régions du visage. Cette connaissance est utilisée pour définir une méthode de reconnaissance se concentrant uniquement sur certaines régions perceptuellement attrayantes du visage et ainsi réduire par un facteur de deux les temps de calcul. Nos résultats montrent une précision de reconnaissance automatique de trois expressions de 99.5% sur la base de données d'expression faciale Cohn-Kanade.

Mots Clef

Reconnaissance des expressions faciales, du visage des régions saillantes, expérimenter eye tracking.

Abstract

Humans can recognize facial expressions in real-time with minimal effort. However, reliable recognition of facial expressions in real-time is a challenging task for computer vision. We present a novel framework that can recognize facial expressions very efficiently and with high accuracy. We propose to computationally process small region of face to extract features. This proposition is based on a psycho-visual experimental study conducted with an eye-tracker. Eye movements of fifteen subjects were recorded with an eye-tracker in free viewing conditions as they watch a collection of 54 videos showing six universal facial expressions. The results of an experimental study show that for some expressions only one facial region is perceptually more attractive than others. While other cases show the attractiveness of two to three facial regions. This knowledge is used in the proposed framework to extract features only from perceptually attractive regions of the face and thus reduction in computational time for feature extraction and in feature vector dimensionality is achieved. Proposed framework achieved automatic expression recognition accuracy of 99.5% for three expressions using Cohn-Kanade facial expression database.

Keywords

Facial expression recognition, facial salient regions, eye tracking experiment.

1 Introduction

Communication in any form i.e. verbal or non-verbal is vital to complete various routine tasks and plays a significant role in daily life. Facial expression is the most effective form of non-verbal communication and it provides a clue about emotional state, mindset and intention [1]. Automatic, reliable and real-time facial expression recognition is a challenging problem due to variability in pose, illumination and the way people show expressions across cultures. Human-Computer interactions, social robots, game industry, synthetic face animation, deceit detection, interactive video and behavior monitoring are some of the potential application areas that can benefit from automatic facial expression recognition.

Humans are blessed with the amazing ability to decode facial expressions across different cultures, in diverse conditions and in a very short time. Despite its limited neural resources the human visual system (HVS) is able to rapidly analyze complex scenes [2]. As an explanation for such performance, it has been proposed that only some visual inputs are selected by considering "salient regions" [3], where "salient" means most noticeable or most important.

In computer vision the notion of saliency was mainly popularized by Tsotsos et al. [4] with their work on visual attention, and by Itti et al. [2] with their work on rapid scene analysis.

In this paper, we propose very efficient and simple method for automatic facial expression recognition. We show that expressions can be recognized very efficiently by mimicking HVS and thus processing only salient regions of face. To determine which facial region(s) is the most important or salient according to HVS, we conducted a psycho-visual experiment using an eye-tracker. We have considered six universal facial expressions for psycho-visual experimental study as these expressions are proved to be consistent across cultures [5]. These six expressions are anger, disgust, fear, happiness, sadness and surprise.

2 State of the art

Feature selection is one of the most important step to successfully analyze and recognize facial expressions automatically. Secondly, it is also very important to extract features only from those region(s) of face that contains discriminative information. 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 [6].

One of the widely studied method to extract appearance information is based on Gabor wavelets [7, 8, 9]. Littlewort et al. [7] has shown a high recognition accuracy (97% for Cohn-Kanade facial expression database [10]) for facial expressions using Gabor features. They proposed to extract Gabor features from the whole face and then selected subset of those features using AdaBoost method. AdaBoost was used to select subset of the features. Tian [8] 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. Donato et al. [9] has employed the technique of dividing the facial image into two : upper and lower face to extract finer Gabor representation for classification. 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. Recently, texture description and classification methods i.e. Local Binary Pattern (LBP) [11] and Local Phase Quantization (LPQ) [12] are also studied to extract appearance-based facial features. Zhao et al. [13] proposed to model texture using volume local binary patterns (VLBP) an extension to LBP, for expression recognition. The authors have proposed to use only the co-occurrences of local binary patterns on three orthogonal planes (LBP-TOP) in order to enhance the applicability of method by reducing the computational complexity. Average facial expression recognition accuracy of 96.26% was achieved for six universal expression with their proposed model using Cohn-Kanade facial expression database. Liao et al. [14] proposed to use two sets of features for expression classification. The first set was obtained by LBP and the second set of features was obtained by linear discriminant analysis (LDA). They tested their model on JAFFE database [15] and achieved average recognition accuracy of 94.59% for seven facial expressions (six universal and one neutral expression). Jiang et al. [16] extended the idea of LBP-TOP to LPQ-TOP and showed that the performance of LPQ based system is better than LBP based system.

For geometric feature-based methods [17, 18, 19], shapes and locations of facial components are extracted to form a feature vector. For expression recognition, Zhang et al. [17] has measured and tracked the facial motion using Kalman Filters. To achieve the recognition task they have also modeled temporal behaviors of facial expressions using Dynamic Bayesian networks (DBNs). In [18] authors have presented Action Unit (AU) detection scheme by classifying features, calculated from "Particle Filter" tracked fiducial facial points. They trained system on the MMI-Facial expression database [20] and tested on the Cohn-Kanade database [10] and achieved recognition rate of 84%. Bai et al. [19] extracted only shape information using Pyramid Histogram of Orientation Gradients (PHOG) and showed the "smile" detection accuracy as high as 96.7% using Cohn-Kanade database. Research has been done with success in recent times to combine features extracted using appearance-based methods and geometric feature-based methods [21, 22].

We have found one shortcoming in all of the reviewed methods for automatic facial expression recognition 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 face 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 study is twofold :

a. We have tried to statistically determine 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 fixation data.

b. We show that very high facial expression recognition (FER) accuracy could be achieved by processing sa-

lient region of face. We propose to extract features only from the salient facial region(s) using Pyramid Histogram of Orientation Gradients [23]. Novel framework for FER achieved recognition accuracy as high as 99.5% using Cohn-Kanade database [10]. The benefit of extracting features only from the salient facial regions is that the framework can be used for real-time applications. Proposed framework processes 4 fps (frames per second) using Matlab 7.6 on Windows PC with 1.8 GHz processor and 1GB RAM. The same machine processes 2fps to extract the same features from the whole face image.

The rest of the paper is organized as follows : all the details related to psycho-visual experiment is described in the next section. Results obtained from psycho-visual experiment and analysis of the data are presented in section 4. Section 5 presents the novel framework for automatic expression recognition and its results on classical database. This is followed by conclusion.

3 Psycho-Visual experiment

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

3.1 Participants, apparatus and stimuli

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

A video based eye-tracker (Eyelink II system, SR Research) was used to record eye movements. The system is capable of compensating head motions. Stimuli were presented on a 19 inch CRT monitor with a resolution of 1024 x 768 and a refresh rate of 85 Hz. A viewing distance of 70 cm was maintained, resulting in a 29° x 22° usable field of view as done by Jost et al. [24].

For the experiment, we used Cohn-Kanade database [10]. 54 videos were selected with the criteria that videos should show both male and female actors, experiment session should complete within 20 minutes and posed facial expression should not look unnatural. Each video (without sound) showed a neutral face at the beginning and then gradually developed into one of the six facial expression. Figure 1 shows example of universal expressions with maximum intensity.

3.2 Eye movement recording

Eye position was tracked at 500 Hz with an average noise less than 0.01°. 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 the observers start viewing videos 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. In fact, severe restrictions in head movements have been shown to alter eye movements and can lead to noisy data acquisition and corrupted results [25].

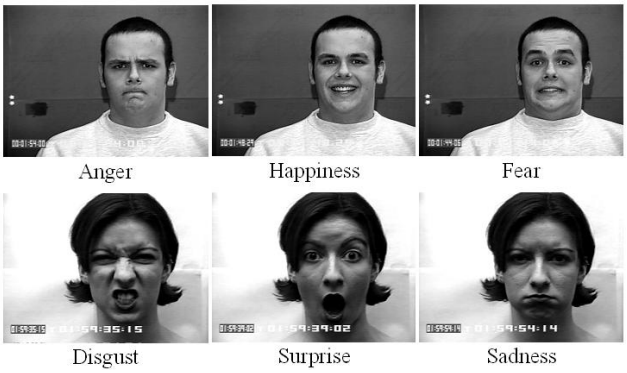

FIGURE 1 – Example of six universal facial expressions from Cohn-Kanade database [10]. Figure showing Peak expression (apex) frame.

4 Psycho-Visual experiment : Results and discussion

In order to know which facial region is perceptually more attractive for specific expression, we have calculated the average percentage of trial time observers have spent on gazing different facial regions. Data are plotted in Figure 2. 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 the video) for analysis of the data. Such misinterpretation could lead to biased data analysis. To meaningfully observe and analyze the gaze trend across one video sequence we have divided each video sequence in three mutually exclusive time periods. The first time period correspond to initial frames of video sequence where the person's face has no emotions i.e. neutral face. The last time period encapsulates the frames where person 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 neutral face to desired expression. Then the fixations recorded for a particular time period are averaged across 15 observers.

Figure 2 shows that the region of mouth is the salient region for the facial expressions of happiness and surprise.

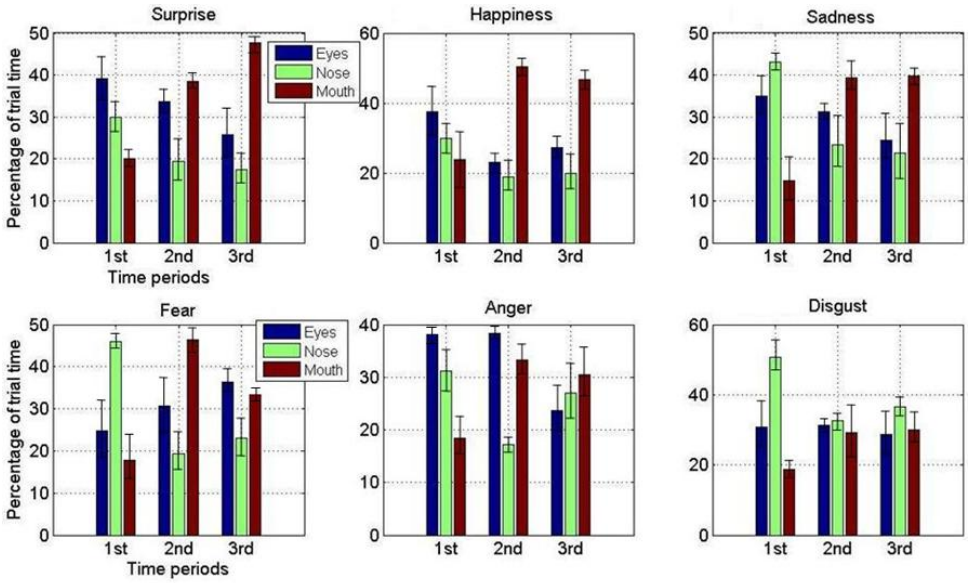

FIGURE 2 – Time period wise average percentage of trial time observers have spent on gazing different facial regions. The error bars represent the standard error (SE) of the mean. First time period : initial frames of video sequence. Third time period : apex frames. Second time period : frames which has a transition from neutral face to particular expression.

This result is consistent with the results shown by Cunningham et al. [26], and Boucher et al. [27]. 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 quite random behavior. Figure 2 demonstrates that even when stimuli show expression with maximum intensity (third time period), observers have gazed all the three regions randomly. One observation that can be made for the expression of disgust is that there is a bit more attraction towards the nose region in the third time period. This could be due to the fact that the wrinkles on the nose region become prominent and attract more attention.

In the expression of fear, facial regions of mouth and eyes attract most of the gazes. From Figure 2 it can be seen that in second time period observers mostly gazed at the mouth region and in the final time period eyes and mouth regions attract most of the attention.

In 1975 Boucher et al. [27] 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 region without any specific trend.

5 Framework for automatic Facial Expression Recognition (FER)

Feature selection along with the region(s) from where these features are going to be extracted is one of the most important step to successfully analyze and recognize facial expressions automatically. In our proposed framework for automatic expression recognition, we extracted Pyramid Histogram of Orientation Gradients (PHOG) [23] features only from the perceptually salient regions. We extracted PHOG features as they have shown to be highly discriminative for FER task [19, 22]. PHOG is a spatial shape descriptor and got its inspiration from the works of Dalal et al. [28] on histograms of oriented gradients and Lazebnik et al. [29] on spatial pyramid matching.

5.1 Feature extraction using PHOG

To test proposed framework for FER we have considered only those expressions out of six universal expressions [5] that have one perceptual salient region. From the analysis of the data from experiment (see Section 4) it has been found that the facial region of "mouth" emerges as the salient region for the expressions of happiness, sadness and surprise. To automatically recognize these expressions PHOG features [23] are extracted only from the mouth region. These features are later used for training classifiers for FER. Steps for the PHOG feature extraction are as follows :

a. Canny edge operator is applied to extract contours from the given stimuli. As illustrated in Figure 3 second row, edge contours represents the shape information.

b. Then image is divided into finer spatial grids by iteratively doubling the number of division in each dimension. It can be observed from Figure 3 that the grid at level l has 2^l cells along each dimension.

c. Afterwards, histogram of orientation (HOG) gradients are calculated using 3 x 3 Sobel mask without Gaussian smoothing and the contribution of each edge is

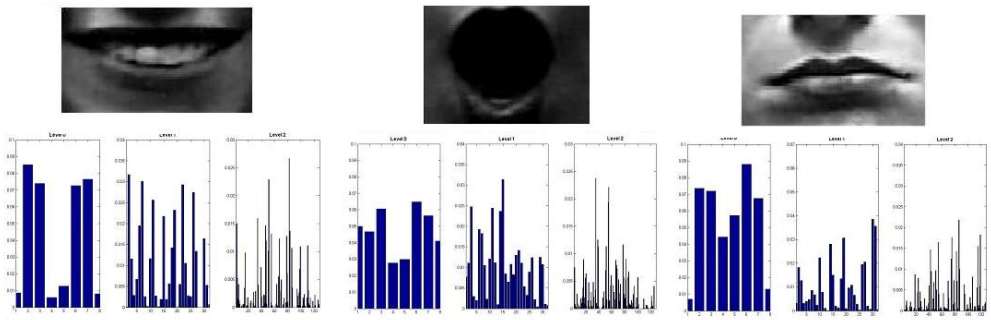

FIGURE 4 – HOG features for different expressions. First column : happiness, second column : surprise, third column sadness. First row shows stimuli and second row shows respective HOG at three levels.

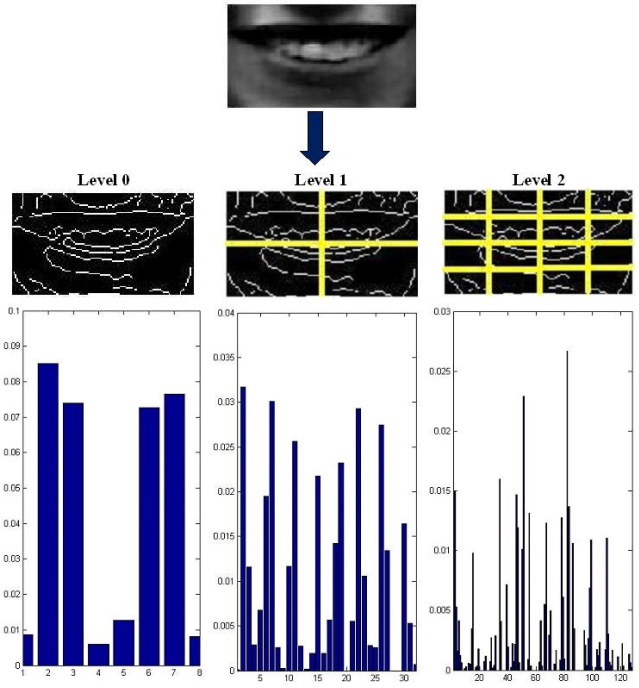

FIGURE 3 – HOG feature extraction. First row : input stimuli, second row : edge contours at three different pyramid levels, third row : histograms of gradients (HOG) at three respective levels.

weighted according to its magnitude. Within each cell, histogram is quantized into N bins. Each bin represents the accumulation of number of edge orientations within a certain angular range.

d. To obtain final PHOG descriptor, histograms of gradients (HOG) at the same levels are concatenated. Final PHOG descriptor is a concatenation of HOG at different pyramid level. Generally, the dimensionality of PHOG descriptor can be calculated by : $N \sum_l 4^l$. In our experiment we obtained 168 dimensional feature vector as we created two pyramid levels with 8 bins with the range of [0-360]. The same is shown in the Figure 3.

Figure 4 shows PHOG features for the expressions of happiness, sadness and surprise. We can observe that PHOG features for these expressions have a discriminative trend.

5.2 Experiments and results

We performed two experiments to measure the performance of proposed framework. For the first study, we use samples from the Cohn-Kanade database [10] and trained three classifiers separately to evaluate the performance of extracted features. Classification accuracy is measured using 10-fold cross validation method. Second experiment evaluates how well the system generalizes on new data. According to our knowledge only Valstar et al. [18] have reported such data earlier.

In the first experiment "Support vector machine (SVM)" with χ^2 kernel and γ=1, "C4.5 Decision Tree (C4.5 DT)" with reduced-error pruning and "Random Forest (RF)" of 10 trees are trained on extracted PHOG features. These parameters are determined empirically. Confusion matrices and recognition accuracy are calculated using 10-fold cross validation. To train and test classifiers we use the same video sequences which we have selected for psycho-visual experiment (see Section 3). We discarded 40% of initial frames from all of the selected videos as the initial frames in Cohn-Kanade (CK) database show neutral expression. After discarding those initial frames we obtain 1012 frames showing one of the three expressions under study. Region of interest (ROI) for all the frames are manually marked and processed to obtain PHOG feature vector. Alernately, ROI (in our case it was mouth region) for the frames can be marked automatically by detecting face from the frames using [30] and then further processing detected face image to detect mouth region using [31]. Average recognition rate of 97.3%, 97.6% and 99.5% are recorded by using SVM, C4.5 Decision Tree and Random Forest respectively. Table 1, 2 and 3 show the confusion matrix for the three classifiers. Diagonal and off-diagonal entries of confusion matrices show the percentages of correctly classified and misclassified samples respectively. Three tables show that the choice of classifier does not effect significantly the recognition accuracy and the extracted PHOG features from the mouth region have high discrimination ability.

	Sadness	Happiness	Surprise
Sadness	100%	0	0
Happiness	0	98.4%	1.6%
Surprise	3.6%	2.8%	93.6%

TABLE 1 – Confusion Matrix : SVM

	Sadness	Happiness	Surprise
Sadness	98.8%	0	1.2%
Happiness	0	97.8%	2.2%
Surprise	0.8%	2.8%	96.4%

TABLE 2 – Confusion Matrix : C4.5 Decision Tree

	Sadness	Happiness	Surprise
Sadness	100%	0	0
Happiness	0	99%	1%
Surprise	0.4%	0	99.6%

TABLE 3 – Confusion Matrix : Random Forest

Aim of the second experiment is to study how well the framework generalizes on new dataset. Thus, the study helps to understand how the different classifiers (same parameters as the first experiment) will behave when they will be used to classify expressions in real life videos. Samples from the CK database (same as the first experiment) are used for the training phase of different classifiers and the samples from FEED database [32] are used for testing. So, the classifiers are tested on samples which are different from training samples altogether. We have chosen FEED database to roughly estimate the accuracy of our framework in real life scenario (with some restrictions), as the CK database exhibits number of drawbacks [18]. Average recognition accuracies using 10-fold cross validation method (for test samples) for the second experiment are presented in Table 4.

	SVM	C4.5 DT	RF
Training samples	97.3%	97.6%	99.5%
Test samples	77.5%	63.2%	79%

TABLE 4 – Average recognition accuracy : training classifier on CK database and testing it with FEED database

Second experiment provided a rough estimate of how accurate the framework will perform in a challenging real life scenario.

6 Conclusion

We conducted a psycho-visual experiment to study and understand how human visual system (HVS) decodes and perceives six universal facial expressions. Eye movements of fifteen observers were recorded using an eye-tracker as they watched the collection of videos showing six universal facial expressions. The analysis of data revealed the fact that visual attention is mostly grabbed by three facial regions i.e. eyes, mouth and nose regions. Conclusions drawn from the experimental study is summarized in Table 5.

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.

TABLE 5 – Summary of the facial regions that emerged as salient for six universal expressions

Secondly, in this paper we presented a novel framework for automatic and reliable facial expression recognition (FER). Framework is based on initial study of human vision. With proposed framework high recognition accuracy (99.5% for CK database), reduction in feature vector dimensionality and reduction in computational time for feature extraction is achieved by processing only perceptually salient region of face. Thus, the framework can be used for real-time application as 4fps (frames/second) can be processed as opposed to 2fps if same features are extracted from whole face image on a same machine. The processing speed can be increased significantly if the framework is implemented using C++ IDE. Another important finding of our study is that the different classifiers trained on extracted PHOG features from "mouth" region are robust enough to perform adequately if they will be used to classify expressions in real videos.

In the future, we will extend proposed framework so that it can recognize wide array of expressions. We will focus on incorporating movement information in our descriptor to make it more robust. Research is required to be done to recognize expressions across camera angle variations.

7 Acknowledgment

This project is funded by the Région Rhône-Alpes, France.

Références

[1] P. Ekman. *Telling Lies : Clues to Deceit in the Marketplace, Politics, and Marriage*. W. W. Norton & Company, New York, 3rd edition, 2001.

[2] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 20, pages 1254–1259, 1998.

[3] L. Zhaoping. Theoretical understanding of the early visual processes by data compression and data selection. *Network : computation in neural systems*, 17 :301–334, 2006.

[4] J. K. Tsotsos, M. C. Scan, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78 :507–545, 1995.

[5] P. Ekman. Universals and cultural differences in facial expressions of emotion. In *Nebraska Symposium on Motivation*, pages 207–283. Lincoln University of Nebraska Press, 1971.

[6] Y. Tian, T. Kanade, and J. F. Cohn. *Handbook of Face Recognition*. Springer, 2005 (Chapter 11. Facial Expression Analysis).

[7] 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.

[8] Y. Tian. Evaluation of face resolution for expression analysis. In *Computer Vision and Pattern Recognition Workshop*, 2004.

[9] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21 :974–989, 1999.

[10] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Fourth IEEE International Conference on Automatic face and Gesture Recognition (FG'00)*, pages 46–53, 2000.

[11] 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.

[12] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *International conference on Image and Signal Processing*, 2008.

[13] 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.

[14] S. Liao, W. Fan, A. C. S. Chung, and D. Yeung. Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In *IEEE International Conference on Image Processing*, 2006.

[15] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 :1357–1362, 1999.

[16] B. Jiang, M. F. Valstar, and M. Pantic. Action unit detection using sparse appearance descriptors in space-time video volumes. In *IEEE Conference on Automatic Face and Gesture Recognition*, 2011.

[17] 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.

[18] M.F. Valstar, I. Patras, and M. Pantic. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 76–84, 2005.

[19] Y. Bai, L. Guo, L. Jin, and Q. Huang. A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition. In *International Conference on Image Processing*, 2009.

[20] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *IEEE International Conference on Multimedia and Expo*, 2005.

[21] 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.

[22] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon. Emotion recognition using PHOG and LPQ features. In *IEEE Automatic Face and Gesture Recognition Conference FG2011, Workshop on Facial Expression Recognition and Analysis Challenge FERA*, 2011.

[23] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *6th ACM International Conference on Image and Video Retrieval*, pages 401–408, 2007.

[24] 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.

[25] H. Collewijin, M. R. Steinman, J. C. Erkelens, Z. Pizlo, and J. Steen. *The Head-Neck Sensory Motor System*. Oxford University Press, 1992.

[26] 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, July 2005.

[27] J. D. Boucher and P. Ekman. Facial areas and emotional information. *Journal of communication*, 25 :21–29, 1975.

[28] N. Dalal, , and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference Computer Vision and Pattern Recognition*, 2005.

[29] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features : spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[30] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[31] L. Cappelletta and N. Harte. Nostril detection for robust mouth tracking. In *Irish Signals and Systems Conference*, 2011.

[32] F. Wallhoff. Facial expressions and emotion database. www.mmk.ei.tum.de/˜waf/fgnet/feedtum.html, 2006.