Facial Expression Recognition using Entropy and Brightness Features

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

[1] LIRIS, UMR 5205 CNRS, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France
[2] Laboratoire Hubert Curien, UMR 5516 CNRS, Université Jean Monnet, Saint-Etienne, France

{Rizwan-Ahmed.Khan, Alexandre.Meyer, Saida.Bouakaz}@liris.cnrs.fr, Hubert.Konik@univ-st-etienne.fr

***Abstract*—This paper proposes a novel framework for universal facial expression recognition. The framework is based on two sets of features extracted from the face image: entropy and brightness. First, saliency maps are obtained by state-of-the-art saliency detection algorithm i.e. "frequency-tuned salient region detection". Then only localized salient facial regions from saliency maps are processed to extract entropy and brightness features. 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.**

***Keywords*-Facial expression recognition, human vision, eye-tracker, salient region detection, entropy, brightness**

I. 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]. Human-Computer interactions, social robots, game industry, synthetic face animation, deceit detection and behavior monitoring are some of the potential application areas that can benefit from automatic facial expression recognition.

Human visual system, despite its limited neural resources has the ability to decode facial expressions in real-time across different cultures and in diverse conditions. As an explanation for such a performance, it has been proposed that only some visual inputs are selected by considering "salient regions" [2], where "salient" means most noticeable or most important. But for computer vision it is very challenging to recognize facial expressions accurately in real-time due to variability in pose, illumination, camera zoom and the way people show expressions across cultures.

Recently, one of the widely studied method for expression recognition task is based on Gabor wavelets [3], [4], [5]. Littlewort et al. [3] 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 [4] 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. [5] 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.

In this paper, we propose a novel framework to efficiently recognize six universal facial expressions [6]. These six facial expressions are anger, disgust, fear, happiness, sadness and surprise. 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 of face that contain discriminative information. In order to determine that which facial region(s) 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 the fixations [7]. 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 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* [8] to localize salient regions of the face. Then localized salient regions are processed to extract two sets of features: entropy and brightness. These two features have been selected for the classification of expressions as

they have shown discriminative abilities.

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

II. Visual experiment

The aim of our 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.

A. 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.

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 viewing distance of 70cm was maintained resulting in a 29° x 22° usable field of view as done by Jost et al. [9].

2) Stimuli: We have used Cohn-Kanade facial expression database [10] which contains 593 videos across 123 subjects. For the experiment we have selected only 54 videos 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 any sound) shows a neutral face at the beginning and then gradually develops into one of the six universal facial expression. Figure 1 shows example of universal expressions with maximum intensity.

B. 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.

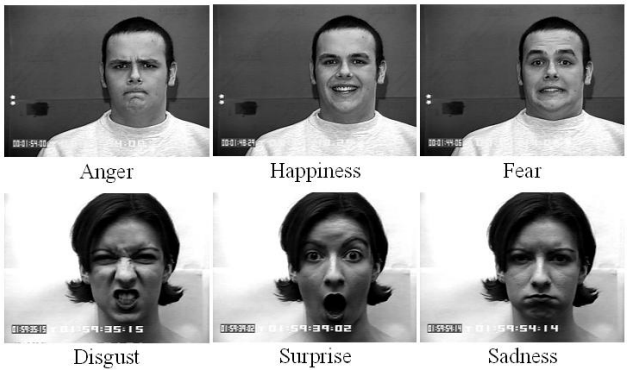

Figure 1. Example of six universal facial expressions from Cohn-Kanade database [10]. Figure showing Peak expression (apex) frame.

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 [11].

III. Results and Discussion : Visual Experiment

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

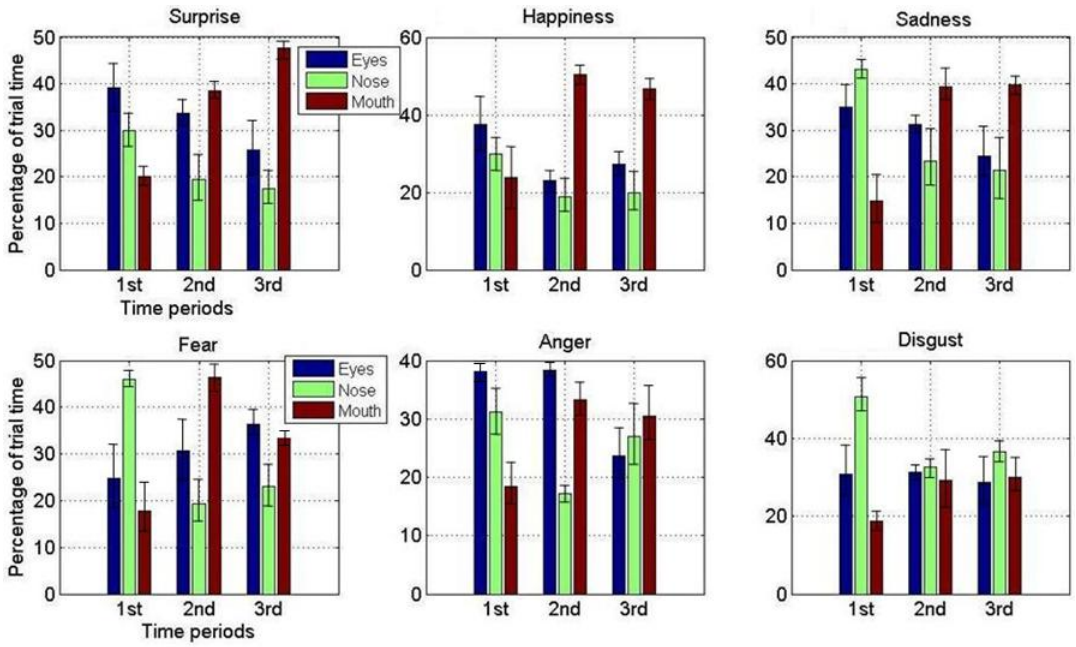

Figure 2. Time period wise average percentage of trial time observers have spent on gazing different facial regions. The error bars represent the standard error (SE) of the mean. First time period: initial frames of video sequence. Third time period: apex frames. Second time period: frames which has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

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. This result is consistent with the results shown by Cunningham et al. [12], and Boucher et al. [13]. 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 intensity, observers have gazed all the three regions randomly (see Figure 2, third time period).

In expression of fear facial regions of mouth and eye attract most of the gazes. From Figure 2 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 and in the final trial period eye and mouth regions attract most of the attention.

In 1975 Boucher et al. [13] 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.

IV. 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 two sets of features only from the salient regions of face. These two features are entropy and brightness.

Entropy is a measure of uncertainty or measure of absence of the information associated with the data [14] while brightness as described by Wyszecki and Stiles [15] 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. In this paper, we have used BCH (Brightness, Chroma, Hue) model [16] for brightness calculation. The first step of proposed framework is to detect salient facial regions.

A. Salient region detection

We propose to use *frequency-tuned salient region detection* algorithm (will be referred as FT later in this paper) developed by Achanta et al. [8] for detection and localization of salient facial regions. This model extracts low-level, pre-attentive bottom-up saliency using center-surround contrast,

color and luminance features. We have chosen this model over other existing state-of-the-art models [17], [18], [19] because it performs better in predicting human fixations (see Figure 3). Secondly, this model is computationally efficient, and is able to extract well-defined salient regions in full resolution. Figure 3 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 (See Figure 3).

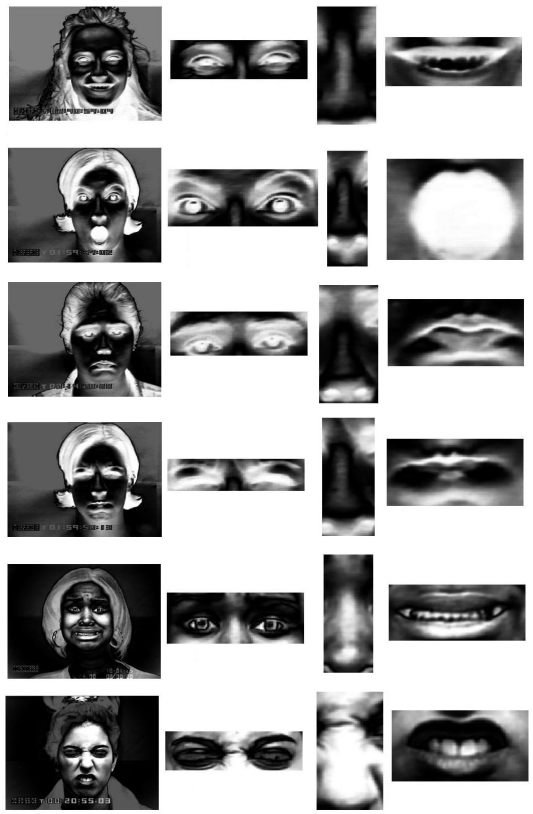

Figure 3. Salient region detection for different expressions using FT [8]. Each row shows detected salient regions in a complete frame along with the zoom of three facial regions i.e. eyes, nose and mouth. Brightness of the salient regions is proportional to its saliency. First row shows expression of happiness, second row: surprise, third row: sadness, fourth row: anger, fifth row: fear and sixth row: disgust .

B. Feature Extraction

We have chosen to extract the features of brightness and entropy for automatic recognition of expressions as these features have shown a discriminative trend, which can be observed from Figure 3 and 4. Figure 4 shows the average entropy values for the facial regions. Each video is divided in three equal time periods for the reasons discussed earlier. 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 4.

By using FT saliency model, we extracted salient regions and obtained saliency maps for every frame of the video (we have used the same 54 videos which were used in visual experiment, see section II). Then obtained saliency maps are further processed for the calculation of brightness and entropy value for the three facial regions. The brightness values, as explained earlier, are calculated using BCH (Brightness, Chroma, Hue) model [16]. The entropy for different facial regions are calculated using equation (1):

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

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 paper 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.

From Figure 4 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 3) 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

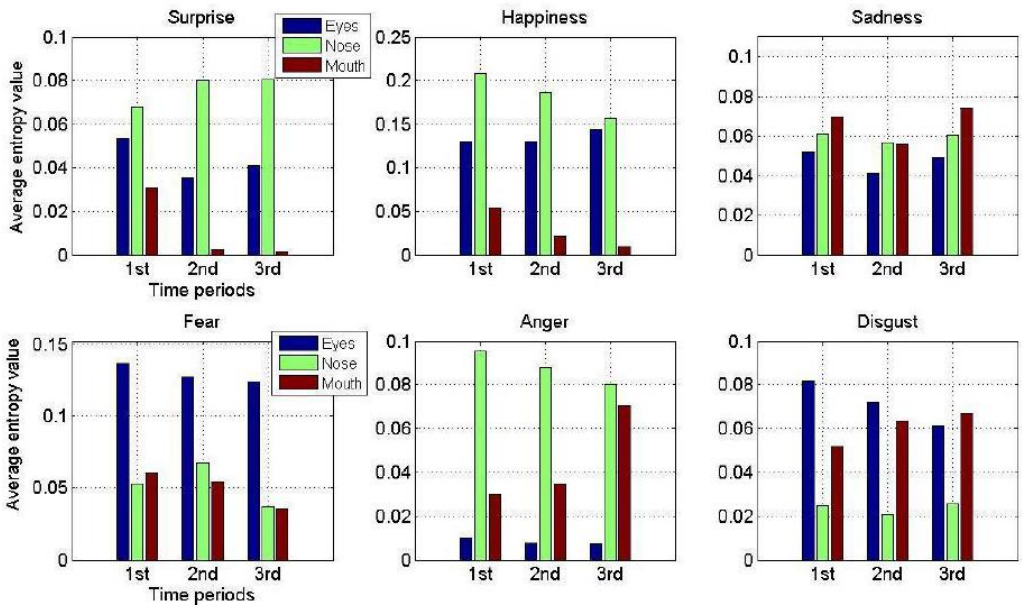

Figure 4. Average entropy value for different facial regions. First time period: initial frames of video sequence. Third time period: apex frames. Second time period: frames which has a transition from neutral face to particular expression.

	Sadness	Happiness	Surprise
Sadness	75.5%	12.8%	11.7%
Happiness	7.2	78.4%	14.4%
Surprise	5.6%	12.8%	81.6%

Table I
CONFUSION MATRIX FOR THE FIRST EXPERIMENT.

	Anger	Disgust	Fear
Anger	71%	15%	14%
Disgust	10.9%	69.3%	19.8%
Fear	13.8%	13.9%	72.3%

Table II
CONFUSION MATRIX FOR THE SECOND EXPERIMENT.

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.

C. Classification results

To measure the performance of proposed approach for facial expression recognition we have completed two experiments. In the first experiment we have considered only those expressions that have found to be decoded and recognized by only one salient region by human visual system. These expressions are happiness, surprise and sadness. In the second experiment the other three facial expressions are considered.

In the first experiment, support vector machine (SVM) with χ^2 kernel and γ=0.5 (parameters are calculated empirically), is trained on the features (entropy and brightness) extracted only from the mouth region of saliency maps. To mimic human visual system, only mouth region is processed to extract features. Confusion matrix and recognition accuracy is calculated using 10-fold cross validation. To train and test classifier we used same video sequences which we have selected for the visual experiment (see Section II). We discarded 40% of initial frames from all of the selected videos as the initial frames in Cohn-Kanade (CK) database [10] show no or neutral expression. After discarding those initial frames we obtain 1012 frames showing one of the three expressions under study. Average recognition rate of 78.5% is recorded for the three expressions. Table I shows the confusion matrix for the first experiment. Diagonal and off-diagonal entries of confusion matrix shows the percentages of correctly classified and misclassified samples respectively.

In the second experiment we trained SVM (with the same parameters as first experiment) on the features selected from the three facial regions i.e. eyes, nose and mouth. The aim in this experiment is to automatically recognize expressions of anger, fear and disgust. In this experiment again, we discarded 40% initial frames of the selected videos showing one of three expressions under study. After discarding initial frames we obtained 858 frames showing expressions of anger, fear and disgust. Average recognition rate of 70.8% is achieved using 10-fold cross validation method. Table II shows the confusion matrix for the second experiment.

Table I and II show that the proposed framework performed adequately in classifying facial expressions by imi-

tating human visual system. The benefit of using proposed framework is that it reduces computational time for feature extraction and thus can be used for real-time applications. The framework processes 5fps (frames/second) as opposed to 2fps if same features are extracted from the whole face image using the same machine (results presented here are obtained from the Matlab's implementation of the framework).

V. Conclusion

The experimental study presented in this paper 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 the six universal 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 the facial expressions can be recognized algorithmically much more efficiently by imitating human visual system. Proposed framework utilizes very well know saliency detection model along with the measure of entropy and brightness. According to our knowledge no scientist has exploited the measure of brightness and entropy to recognize facial expressions. With proposed framework acceptable recognition accuracy, reduction in feature vector dimensionality and reduction in computational time for feature extraction is achieved by processing only perceptually salient region of face.

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 accurate and robust. Research is required to be done to recognize expressions across camera angle variations.

References

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

[3] 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.

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

[5] 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.

[6] 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.

[7] U. Rajashekar, L. K. Cormack, and A.C Bovik. Visual search: Structure from noise. In *Eye Tracking Research & Applications Symposium*, pages 119–123, 2002.

[8] Achanta R., Hemami S., Estrada F., and Susstrunk S. Frequency-tuned salient region detection. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2009.

[9] 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.

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

[12] 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.

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

[14] C. E. Shannon and W. Weave. *The Mathematical Theory of Communication*. University of Illinois Press, 1963.

[15] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley-Interscience, 2000.

[16] S. Bezryadin and P. Bourov. Color coordinate system for accurate color image editing software. In *International Conference Printing Technology*, pages 145–148, 2006.

[17] 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.

[18] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[19] C. Koch J. Harel and P. Perona. Graph-based visual saliency. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2006.