LOCAL SALIENCY-INSPIRED BINARY PATTERNS FOR AUTOMATIC RECOGNITION OF MULTI-VIEW FACIAL EXPRESSION

Bikash Santra, Dipti Prasad Mukherjee

Indian Statistical Institute, 203, B. T. Road, Kolkata 700108, INDIA

ABSTRACT

In this paper, we propose a novel scheme for automatic recognition of facial expressions captured from both fronto-parallel and non-fronto-parallel cameras i.e., multi-view facial expressions (MVFE). The proposed scheme introduce a *Local Saliency-inspired Binary Pattern (LSiBP)* feature to recognize MVFE. First view-specific approximated saliency likelihood map (ASLM) is derived during training of our model. ASLM is determined from the 2D structure tensor representation of faces. The distribution of saliency likelihoods of pixels along with pixel intensities are analyzed in extracting *LSiBP* features from a face. Such *LSiBP* features are utilized for training and testing of view-specific SVM classifiers. Extensive experiments are performed on datasets of both posed and unconstrained spontaneous expressions of MVFE. Our scheme outperforms state-of-the-arts by at least 1% to recognize MVFE.

Index Terms— Local Saliency-inspired Binary Pattern (LSiBP), Local Binary Pattern (LBP), multi-view facial expression, KDEF, SFEW, LFPW, structure tensor, saliency likelihoods.

1. INTRODUCTION

Facial expression manifests the affective state of mind and it has a major impact in non-verbal communication of human beings [1]. An automatic facial expression recognition system can be utilized in many application areas like human-computer interaction, human affective behavior analysis, *etc.* In last two decades, a remarkable progress has been accomplished in recognizing emotional expressions from near frontal images or image sequences [2–4]. But, recognition of multi-view facial expression (MVFE) is still an unexplored research area compare to recognition of expressions from near frontal images.

The accurate recognition of MVFE from 2D face images is a challenge using a computationally inexpensive model. This is due to occlusion of part of face in 2D images, and coupling of views and emotions. In this paper, we accept those challenges of analyzing facial behaviour from 2D face images. In faces, the expression related information lies only on a specific salient regions of the faces. Motivated by the fact, we propose a spatial saliency inspired feature extraction scheme for recognition of MVFE.

Recent advances in image acquisition technology from 2D to 3D images have facilitated new approaches [5–9] for recognition of MVFE. These methods for recognition of MVFE can be grouped into two categories: 1) Geometry based methods [5], and 2) Appearance based methods [6–9]. Geometry based methods require accurate localization of landmark points on faces while facial pixel intensities are the primary concern in appearance based methods.

An example of first group [5] extracts geometric features in capturing 2D displacement of facial landmark points between expressions and *neutral* for recognition of MVFE. In case of [8, 9] the appearance based features local binary patterns (LBP) [10] are extracted around facial landmark points for recognition of MVFE. Similarly, LBP and variants of LBP features are extracted from the faces of MVFE in [7]. In [6], dense scale invariant features (DSIFT) [11] are extracted at a set of facial landmark points followed by dimensionality reduction. These approaches are more or less dependent on facial key points as the features are extracted around facial key points. But, the proper detection of facial key point is still a challenging problem for multi-view faces. So, the erroneous detection of facial key points leads to misclassification of expression. Therefore, we intend to introduce an appearance based *Local Saliency-inspired Binary Pattern (LSiBP)* feature for accurate recognition of MVFE. Till today, there exists no salient region based framework for recognition of MVFE as per our knowledge. In recognition of expression from near frontal faces, [12] utilizes facial saliency to represent expressive faces and [13] identifies facial components through facial saliency. But, we neither represent expressive faces nor extract facial components by salient regions of faces. We utilize the whole facial saliency likelihood map indicating degree of saliency of pixels in extracting features from MVFE. In saliency detection, structure tensor method shows good performance in [14]. Motivated by [14], we utilize a directional coherence measure from the 2D structure tensor representation of faces in detecting facial saliency.

The contribution of this paper as compared to the previous related studies is three-fold. First, we introduce a saliency based framework for recognition of MVFE as per our best knowledge. Second, we introduce a variant of LBP feature extraction scheme *Local Saliency-inspired Binary Patterns (LSiBP)* to automatically recognize MVFE. LBP features are calculated with only pixel intensities of an image. But, *LSiBP* features are computed by observing marginal saliency likelihoods of pixels along with pixel intensities of faces to accurately recognize MVFE. Third, we present a 2D structure tensor representation of faces for the first time in computing facial saliency for recognition of MVFE. The rest of the paper is organized as follows. Proposed methodology is described in Section 2. Experimental results and comparisons are shown in Section 3. Finally, Section 4 concludes the paper.

2. METHODOLOGY

The view angles of MVFE lies in the interval $[-90°, 90°]$. Let, θ be the view angles and $\Delta\theta$ be the interval between two consecutive views. In [15], Zhu *et al.* detect landmark points and estimate the view angle in $[-90°, 90°]$, where $\Delta\theta = 15°$. [15] provides 68 landmark points for faces in $[-45°, 45°]$. In our model, we take θ in $[-45°, 45°]$. At each θ, the set of 68 landmark points of a face image of emotion *neutral* is taken as canonical frame. Subsequently, all the face images at each θ are aligned to the canonical frame by calculating an affine transformation. The face regions are extracted from aligned faces and converted to the gray images. Now, we have the pre-processed face images for rest of the procedure. The remaining modules of our model are described successively in the following subsections.

2.1. Approximated Saliency Likelihood Map

We determine view-specific approximated saliency likelihood map (ASLM) for MVFE in $[-45°, 45°]$. The process of finding ASLM is discussed here for a particular view $\theta \in [-45°, 45°]$. The ASLM for remaining views can be determined in similar fashion. In face images, only specific salient regions primarily characterize different emotional expressions. For example, disgust is characterized mainly by nose and upper lip movement whereas fear is expressed by movement of upper eyelid. In this work, our motivation is to utilize the distribution of degree of saliency of pixels for MVFE recognition. The proposed scheme has two phases such as train and test. Consecutively, dataset of MVFE at θ view are subdivided into two sets such as training data and test data. Using training data, the ASLM is determined through two consecutive steps. First we compute facial saliency likelihood maps (FSLMs) for all faces from training data by 2D structure representation of faces. Last, a Bayesian inference based fusion of FSLMs are presented for ASLM. The derivation of FSLMs are presented next.

2.1.1. Facial Saliency Likelihood Maps

The saliency map of a face is derived using a 2D structure tensor representation of the face. The 2D structure tensor [16] is designed using the local gradient field along spatial axes. Let, $\mathtt{I}$ be an $M \times N$ face image from training data. We consider a $d \times d$ region R centered at pixel $(x, y) \in \mathtt{I}$ (we set $R = 3 \times 3$ pixels). The 2D structure tensor for pixel (x, y) is defined as:

$$\mathtt{F}(x,y) = \begin{bmatrix} \sum_{(k,l)\in R} \mathtt{I}_X(k,l)^2 & \sum_{(k,l)\in R} \mathtt{I}_X(k,l)\mathtt{I}_Y(k,l) \\ \sum_{(k,l)\in R} \mathtt{I}_X(k,l)\mathtt{I}_Y(k,l) & \sum_{(k,l)\in R} \mathtt{I}_Y(k,l)^2 \end{bmatrix},$$

where $\mathtt{I}_X(k,l)$, and $\mathtt{I}_Y(k,l)$ denote the gradients at (k,l) along X-axis and Y-axis respectively. The eigenvalues of $\mathtt{F}(x,y)$ indicate major directions of gradients in R corresponding to (x,y). In [16], it is shown that the directions corresponding to the largest eigenvalue (say $\lambda_{(x,y)}$) is the dominant direction. Consequently, the saliency map of $\mathtt{I}$ is obtained as $\mathtt{D} = [\lambda_{(x,y)}]$. Now, $\mathtt{D}$ is normalized into the range $[0, 1]$ to obtain facial saliency likelihood map (FSLM) of $\mathtt{I}$. In this way, we determine the FSLMs for all faces of training data. The fusion of FSLMs in deriving ASLM is presented next.

2.1.2. Fusion of Saliency Maps

Assume that we have L number of faces in training data. Correspondingly, we get L number of FSLMs. Let the i-th FSLM is denoted by $\mathtt{D}_i$. Initially the correspondence between FSLMs and images from training data are not known explicitly. Our goal is to find a fused saliency map from these L FSLMs. We do not compute saliency map for test data to optimize computational cost during test. The fused saliency map is constructed such a way that the fused saliency map approximately corresponds to any face image.

For a pixel (x, y) in an image $\mathtt{I}$, the probability of the pixel being salient is denoted by $P_{\mathtt{I}}(x, y)$. We obtain $P_{\mathtt{I}}(x, y)$ by using Bayesian probabilistic framework. The joint probability $P((x,y), \mathtt{D}_i) = P_{\mathtt{I}}((x,y)|\mathtt{D}_i)P_{\mathtt{I}}(\mathtt{D}_i)$, such that: $\sum_{(x,y)} \sum_i P((x,y), \mathtt{D}_i) = 1$. From this model, the marginal saliency likelihood of (x,y) for $\mathtt{I}$ can be estimated by the marginal likelihood distribution $P'_{\mathtt{I}}(x,y)$. Using Bayesian inference rule, we obtain:

$$P'_{\mathtt{I}}(x,y) = \sum_i \underbrace{P_{\mathtt{I}}((x,y)|\mathtt{D}_i)}_{\text{Likelihood}} \underbrace{P_{\mathtt{I}}(\mathtt{D}_i)}_{\text{Prior}}, \tag{1}$$

where the prior $P_{\mathtt{I}}(\mathtt{D}_i)$ is the probability that $\mathtt{D}_i$ is saliency map of $\mathtt{I}$ and $P_{\mathtt{I}}((x,y)|\mathtt{D}_i)$ is the probability that pixel (x,y) of image $\mathtt{I}$ is salient in $\mathtt{D}_i$.

$P_{\mathtt{I}}((x,y)|\mathtt{D}_i)$ is obtained from the saliency map $\mathtt{D}_i$. The challenge lies in deriving the prior $P_{\mathtt{I}}(\mathtt{D}_i)$. In our problem, $P_{\mathtt{I}}(\mathtt{D}_i)$ is associated with the variance of FSLMs. A FSLM with higher variance of saliency likelihood (SL) has more impact on fused saliency map. So, more confidence score is assigned to a FSLM with higher variance of SL as follows: $P_{\mathtt{I}}(\mathtt{D}_i) = \frac{var(\mathtt{D}_i)}{\sum_{j=1}^{L} var(\mathtt{D}_j)}$. Therefore, the marginal likelihood distribution $P'_{\mathtt{I}}(x,y)$ can be computed by providing the likelihood and prior in Eq. (1). The approximated saliency likelihood is $P_{\mathtt{I}}(x,y) = \frac{P'_{\mathtt{I}}(x,y)}{\sum_{(x,y)\in\mathtt{I}} P'_{\mathtt{I}}(x,y)}$. Consequently, we get an approximated saliency likelihood map (ASLM), $\mathtt{S} = [P_{\mathtt{I}}(x, y)]$. Since, $\mathtt{S}$ is constructed using training data, $\mathtt{S}$ contain saliency likelihoods of all possible expressions with variations in illumination and appearance of faces. The *LSiBP* feature extraction is presented next.

2.2. Feature Extraction & Classification

In the proposed scheme, we introduce a variant of LBP termed as *local saliency-inspired binary patterns (LSiBP)*. Let, $\mathtt{I}'$ be a face image of MVFE. First, we select view-specific ASLM $\mathtt{S}$ corresponding to the view of $\mathtt{I}'$. In our formulation, $\mathtt{I}'(x, y)$ and $\mathtt{S}(x, y)$ indicate intensity and saliency likelihood respectively for pixel (x, y) in image $\mathtt{I}'$. *LSiBP* value for pixel (x, y) is derived over a p-neighbourhood of (x, y) in $\mathtt{I}'$ as well as in $\mathtt{S}$. The p-neighbourhood pixels are determined by considering a circle of radius r centered at (x, y). Therefore, a nighbourhood pixel (x_m, y_m) is located at $x_m = x - rsin(2\pi m/p)$ and $y_m = y + rcos(2\pi m/p)$, where $0 \leq m \leq p - 1$. For each (x_m, y_m) in neighbourhood of (x, y), we first find a binary number with the bit b_m for each m as:

$$b_m = \begin{cases} 1 & \text{if, } \mathtt{I}'(x_m, y_m) \geq \mathtt{I}'(x, y), \\ & \quad \text{and } \mathtt{S}(x_m, y_m) \geq \mathtt{S}(x, y) \\ 0 & \text{otherwise.} \end{cases}$$

Then, for pixel (x, y), the *LSiBP* is the p-bit binary number $b_{p-1}...b_1 b_0$ and represented as: $LSiBP_{r,p}(x, y) = \sum_{m=0}^{p} b_m 2^m$.

An uniform pattern encloses maximum 2 bitwise transitions from 0 to 1 (or, from 1 to 0) for the circular representation of a binary number. Hence, there exists a total of $p(p-1)+2$ uniform patterns for p-neighbourhood of a pixel. It is seen that uniform patterns cover most of the different types of LBP patterns that appear in texture representing facial expressions [10]. Like uniform LBP patterns, uniform *LSiBP* patterns significantly reduce the number of feature dimensions and still represent reliable expression-related information. The proposed *LSiBP* is an uniform *LSiBP*.

In facial expression recognition, the whole face is a global representation of facial components for an emotional expression. A feature computed over the whole face image represents only the occurrences of the micro-patterns without indicating their proper locations. Therefore, we need to extract features in which the local information is preserved. In order to achieve it, the region-based feature representation approach is adopted. Hence, the face image is subdivided into several blocks by the number of vertical and horizontal grid lines. After analyzing recognition accuracies in several experiments, we set 9 rows and 8 columns of overlapping blocks with an overlap of 6 pixels. In our scheme, we extract $LSiBP_{3,8}$ feature for each block of a face, where r and p are chosen after number of experiments. Now, we calculate a histogram of *LSiBP* codes for each block of the face and concatenate them into a single histogram in order. This concatenated histogram of the *LSiBP* codes uniquely represent the face. The classification of expression features is presented next.

In our proposed model, we utilize support vector machine (SVM) [17] classifier. Our model is trained and tested with view-specific SVMs for each view angle θ. We select radial basis function (RBF) kernel of SVM for our experiments. Experimental procedure and results are presented next.

3. EXPERIMENTS

A number of experiments are conducted with 3 view angles $-45°, 0°$ and $45°$ and 7 expressions such as anger, disgust, fear, joy, neutral, sadness, surprise. Experiments are extensively performed on freely available posed (KDEF [18]) and spontaneous (LFPW [19] and SFEW [20] datasets in the wild) expression datasets. Descriptions on datasets are follows: (i) The Karolinska Directed Emotional Faces (KDEF) [18] dataset includes total 4900 number 2D images of 7 expressions of 70 subjects. At each view angle, each expression is framed twice for all subjects. We collect 1168 images from KDEF. The selection criteria are: (a) images from first session, (b) images with view angles $-45°$, $0°$, and $45°$, and (c) images in which 68 landmark points detected by [15]. (ii) LFPW dataset [19] includes face images of smile expression having large variations in view angle, illumination and occlusion (downloaded from google.com, yahoo.com, flickr.com, etc.). We have collected total 88 face images of smile expression from the provided live links. We manually classify the images into 3 views $-45°$, $0°$, and $45°$ for our experiments. (iii) The SFEW dataset [20] includes 700 images of 95 subjects. The images are extracted from movies depicting facial expression with different view angles, illumination conditions and occlusions. The images from the dataset are annotated with five landmark points and labeled into 7 expression categories. The images are manually classified into 3 views $-45°$, $0°$, and $45°$ for experiments. Experimental results are described next.

3.1. Experiments on KDEF

First, the pre-processed face images from KDEF scaled to the size of 168×188 and 3 view-specific ASLMs are derived using training data. Consecutively, the $LSiBP_{3,8}$ features are extracted from faces of KDEF. 3 view-specific SVMs are trained and tested with the expression features. Note that, during test, the view angle is determined by [15] and pre-processed as discussed in Section 2. We perform 10 fold subject independent cross-validation. The parameters of the RBF-kernel of SVM are tuned and consequently the results are generated by

Table 1. Recognition accuracy (%) per expression and view on KDEF. AN: Anger, DI: Disgust, FE:Fear, JO: Joy, NE: Neutral, SA: Sadness, SU:Surprise, and AVG: Average.

View Angles	**Expression**							**AVG**
	AN	**DI**	**FE**	**JO**	**NE**	**SA**	**SU**	
$-45°$	72.46	86.57	67.24	**97.10**	86.96	72.16	80.30	80.39
0°	79.69	88.71	66.67	**96.88**	95.38	72.13	89.06	**84.07**
45°	68.09	73.33	55.82	**93.33**	91.23	64.58	40.00	77.03

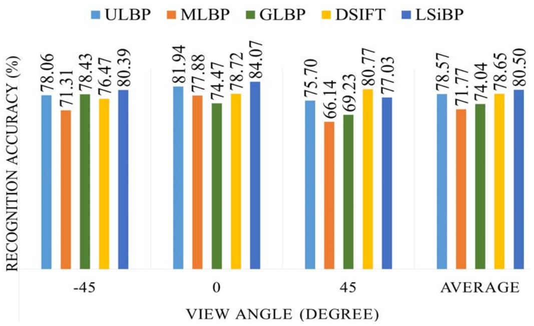

Fig. 1. Comparison on tested features with KDEF dataset.

cross-validation. Table 1 summarizes the recognition accuracy (i.e., total number of correct classification of test images / total number of test images) in percentage. From Table 1, it is seen that the proposed scheme achieves highest recognition accuracy 84.07% at $-45°$ view on average. In each view angle, the emotion *joy* consistently gets highest recognition accuracy. We get low classification rate of our model at $+45°$ view for all the expressions except the emotion *neutral*. A probable reason of it could be the less number of data compare to other views.

Table 2. Recognition accuracy $(\%)$ of smile detection for different features from LFPW by the model trained with KDEF.

Features	**View Angles**			**Average**
	$-45°$	$0°$	$+45°$	
LBP [5, 7–9]	85.71	62.12	90.00	79.28
MLBP [7]	42.86	71.21	80.00	64.69
LGBP [7]	80.95	**100.00**	60.00	80.32
DSIFT [6]	33.33	50.00	50.00	44.44
LSiBP (Proposed)	**90.48**	81.82	**100.00**	**90.76**

We compare the performances of our model with *LSiBP* features with other features used in state-of-the-arts for recognition of MVFE. We select LBP [5, 7–9], magnitude local binary patterns (MLBP) [7], local gabor binary patterns (LGBP) [7] and DSIFT [6] features for comparison as these all shows good performances in recognizing MVFE. In our framework, we extract LBP, MLBP, LGBP and DSIFT features as replacement to *LSiBP* and evaluate the results on KDEF. It is clearly seen in Fig. 1 that *LSiBP* outperforms all the competing features for recognition of MVFE. In the next subsection, the cross dataset experiments are presented.

3.2. Cross Dataset Experiments on KDEF and LFPW

In this experiment the proposed scheme is trained with KDEF dataset and tested with LFPW dataset.The pre-processed faces from KDEF are scaled to the size of 168×188. In case of LFPW dataset, the faces are pre-processed with respect to the canonical frame for each view from KDEF dataset and scaled to the size of 168×188. 3 view-specific ASLMs are derived during the training with images of 7 expressions from KDEF. Consecutively, $LSiBP_{3,8}$ features are extracted from faces of KDEF and LFPW. 7 expression features from KDEF are used to train per-view SVMs. The parameters of the RBF-kernel of SVMs are tuned during training with KDEF. In case of test, expression features of the emotion *smile* (i.e., *joy*) are fed to trained SVMs at the view determined during pre-processing of the face. Similarly, experiments are conducted for LBP, MLBP, LGBP, and DSIFT features as replacement of *LSiBP*. In Table 2, it is seen that the *LSiBP* achieves highest recognition accuaracy 90.76% on average over other features. At $-45°$ and $45°$ view, *LSiBP* shows better performances while LGBP performs better at $0°$.

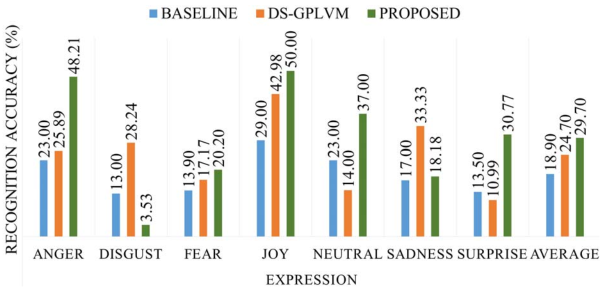

Fig. 2. Recognition accuracy per expression class on different methods trained and tested with the SFEW dataset.

3.3. Experiments on SFEW of Spontaneous Expression

The images of the SFEW dataset are provided into two subject independent sets such as Set-1 and Set-2. All the faces of 7 expressions from SFEW are aligned using five landmark points provided with dataset followed by cropping of face regions. Next, we scale the images to the size of 168×188 and 3 view-specific ASLMs are derived during training. The $LSiBP_{3,8}$ features are extracted from faces of SFEW. Now, we train per-view SVMs with expression features of Set-1 and test with that of Set-2, and vice-versa. The results for different methods are presented in Fig. 2. The 'PROPOSED' in the figure shows results of our model with *LSiBP* features. In Fig. 2, the results obtained by the originator [20] of the dataset are shown by 'BASELINE' and 'DS-GPLVM' present the results of multi-view learning method in [9]. It is clearly observed in Fig. 2 that our method outperforms state-of-the-arts by 5% on average. Also, the proposed method yields highest result for 5 expressions. Next section concludes the proposed work.

4. CONCLUSION

In this paper, a novel approach is proposed to automatically recognize MVFE by introducing saliency based *LSiBP* feature. *LSiBP* feature characterize the emotional expressions by utilizing degree of saliency of pixels for manifestation of human facial expression. The *LSiBP* of expressions are classified by SVMs per-view. The proposed scheme is experimented with one posed and two spontaneous expression dataset. Our scheme outperforms the state-of-the-arts for almost all the cases. But, we need an accurate classification of spontaneous expression from the real world images. Hence, we aim is to extend our work for proper detection of spontaneous expression from the real world images.

5. REFERENCES

[1] Jeffrey F Cohn, "Foundations of human computing: facial expression and emotion," in *Proceedings of the 8th international conference on Multimodal interfaces*. ACM, 2006, pp. 233–238.

[2] Yongqiang Li, Shangfei Wang, Yongping Zhao, and Qiang Ji, "Simultaneous facial feature tracking and facial expression recognition," *Image Processing, IEEE Transactions on*, vol. 22, no. 7, pp. 2559–2573, 2013.

[3] Caifeng Shan, Shaogang Gong, and Peter W McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[4] Guoying Zhao and Matti Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[5] Yuxiao Hu, Zhihong Zeng, Lijun Yin, Xiaozhou Wei, Jilin Tu, and Thomas S Huang, "A study of non-frontal-view facial expressions recognition," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[6] Wenming Zheng, Hao Tang, Zhouchen Lin, and Thomas S Huang, "Emotion recognition from arbitrary view facial images," in *Computer Vision–ECCV 2010*, pp. 490–503. Springer, 2010.

[7] S Moore and R Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541–558, 2011.

[8] Nikolas Hesse, Tobias Gehrig, Hua Gao, and Hazim Kemal Ekenel, "Multi-view facial expression recognition using local appearance features," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 3533–3536.

[9] Stefanos Eleftheriadis, Ognjen Rudovic, and Maja Pantic, "Discriminative shared gaussian processes for multiview and view-invariant facial expression recognition," *Image Processing, IEEE Transactions on*, vol. 24, no. 1, pp. 189–204, 2015.

[10] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[11] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] Sima Taheri, Vishal M Patel, and Rama Chellappa, "Component-based recognition of facesand facial expressions," *Affective Computing, IEEE Transactions on*, vol. 4, no. 4, pp. 360–371, 2013.

[13] SL Happy and Aurobinda Routray, "Automatic facial expression recognition using features of salient facial patches," *Affective Computing, IEEE Transactions on*, vol. 6, no. 1, pp. 1–12, 2015.

[14] Wonjun Kim and Changick Kim, "Spatiotemporal saliency detection using textural contrast and its applications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 4, pp. 646–659, 2014.

[15] Xiangxin Zhu and Deva Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2879–2886.

[16] Hans Knutsson, "Representing local structure using tensors," in *6th Scandinavian Conference on Image Analysis, Oulu, Finland*. Linköping University Electronic Press, 1989, pp. 244–251.

[17] Corinna Cortes and Vladimir Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[18] D Lundqvist, A Flykt, and A hman, *The Karolinska Directed Emotional Faces - KDEF*, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.

[19] Peter N Belhumeur, David W Jacobs, David J Kriegman, and Narendra Kumar, "Localizing parts of faces using a consensus of exemplars," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 12, pp. 2930–2940, 2013.

[20] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2106–2112.