Micro-Expression Recognition using Histogram-Based Feature Descriptors

Kevin Zhang (kzhang@cs.wisc.edu), Satyam Dhar (sdhar@cs.wisc.edu)

Abstract

Human facial expressions contain substantial information about how a person is feeling at any instant. People can fake their facial expressions but micro-expressions, which are short, involuntary facial expressions, can reveal a person's underlying emotions even when the person is trying to conceal their feelings. In this project, we build a system for automatic facial expression recognition (FER) and later extend this system for micro-expression recognition (MER). For FER, we evaluate and compare performance of two different feature descriptors, Local Binary Patterns (LBP) and Local Phase Quantization (LPQ). For MER, we examine the performance of LBP on Three Orthogonal Planes (LBP-TOP) and evaluate the performance of the LPQ-TOP feature descriptor which has not been applied to this problem in the past. Our results show that LPQ and its variants perform slightly better than LBP for both facial expression recognition and micro-expression recognition in our experiments on the Extended Cohn Kanade and CASME II datasets.

Introduction

Facial expressions contain valuable information about a person's emotions. A smile can indicate happiness and a furrowed brow can indicate anger. Humans are very good at recognizing these macro-expressions which occur on the order of seconds but people can also manipulate their faces in order to conceal their true emotions. Micro-expressions (ME) are short, involuntary facial expressions that can reveal a person's underlying emotions even when the person is trying to conceal their feelings [1]. Because MEs last for fractions of a second, even experts trained in state of the art techniques have trouble consistently recognizing MEs correctly [2]. One of the major reasons for widespread interest in MEs is that they provide clues in lie detection [1]. Since MEs are spontaneous and involuntary in nature, subjects in high stress situations, such as police interrogations, can display MEs indicating guilt even if they tell a false story of innocence. MEs can also be used to gauge reactions to advertising campaigns, detect fraud in insurance claims, or reveal a patient's genuine emotions in psychotherapy.

While there have been many computer vision techniques developed for facial expression (FER) recognition, we are interested specifically in ME recognition because it has a unique set of challenges. A simple facial expression recognition system for instance, must discern an image of a smiling face from an image of a frowning face – two images which are obviously different. Since MEs are relatively more subtle than normal FEs, we face the problem of classifying two images which look almost identical at first glance. We believe that ME recognition is an area where human vision falls short. We first evaluate some state of the art techniques for FER using histogram-based descriptors including local binary patterns (LBP) and local phase quantization

(LPQ). We then apply variations of these techniques to the problem of micro-expression recognition.

Previous Work

Facial Expression recognition has received a lot of attention due to its applications in human-computer interaction, data-driven animation, and robotics. There are two common approaches to extract facial features: geometric feature-based methods and appearance-based methods [1]. Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry. Based on their work in detecting Action Units through classification of features that represent tracked facial points, Valstar et al. [4,5] argued that this kind of geometric facial representation is well suited for facial expression analysis. However, geometric feature-based methods usually require accurate facial feature detection and tracking, which is hard to obtain in many situations. Among appearance-based methods, image filters are applied to either the whole-face or specific face-regions to extract the appearance changes of the face. One such commonly used filter is Gabor wavelets. The major works on appearance-based methods have focused on using Gabor-wavelet representations [9,6,7,10,13] because of their better performance and accuracy. One drawback, however, is that applying these filters are time-intensive and also require a lot of memory. For example, in [13], the Gabor-wavelet representation derived from each 48x48 face image has a time complexity of $O(10^5)$.

The ME recognition task attempts to classify a given image sequence that contains a micro-expression into two or more classes. Research in this area has been conducted using both posed and spontaneous data. Some researchers investigated ME recognition on posed ME databases. Polikovsky et al. [18, 19] used a 3D gradient descriptor for the recognition of AU-labeled MEs. Wu et al. [20] combined Gentleboost and an SVM classifier to recognize synthetic ME samples from the METT training tool. Li et al. [21,22] combined three steps: (i) a temporal interpolation model (TIM) to temporally 'expand' the micro-expression into more frames, (ii) LBP-TOP feature extraction (after detecting facial feature points with an Active Shape Model (ASM)); and (iii) Multiple kernel learning (MKL) for classification and evaluate its performance on two challenging spontaneous ME databases (SMIC and CASMEII). LBP and its variants have often been employed as the feature descriptors for ME recognition in many other studies. Ruiz-Hernandez and Pietikainen [23] ¨ used the re-parameterization of a second order Gaussian jet to generate more robust histograms, and achieved better ME recognition result than [21] on the first version of SMIC database (six subjects). Wang et al. [24] extracted LBP-TOP from a Tensor Independent Colour Space (TICS) (instead of ordinary RGB) for ME recognition, and tested their method on CASMEII database. In their another paper [25], Local Spatiotemporal Directional Features (LSDF) were used together with the sparse part of Robust PCA (RPCA) for ME recognition, achieving an accuracy of 65.4% on CASMEII. So far most ME recognition studies have considered using LBP-TOP as the feature descriptor. Since there is scope for improvement in the accuracy for the problem of ME recognition, new techniques need to examined to be used as feature descriptors. In this work, we evaluate the performance of

LBP-TOP as well as LPQ-TOP, a feature descriptor which has not been previously applied to the problem of ME recognition.

Many machine learning techniques have been used in the past for both FER and MER including Neural Networks [16,15,9], Support Vector Machines (SVM) [13], Bayesian Networks (BN) [14] and rule-based classifiers [8, 11, 12]. According to the comparison done by Bartlett et al. [13], best results for FER were obtained by selecting a subset of Gabor filters using AdaBoost and then training SVM on the outputs of the selected filters. Since SVMs have empirically performed the best in facial recognition applications, we implement our system using SVMs.

Techniques for Feature Extraction

In this project, we examine two different techniques to generate feature descriptors for the facial image sequences namely Local Binary Patterns (LBP) and Local Phase Quantization (LPQ) which we describe in more detail in the following paragraphs.

Local Binary Patterns

In order to apply machine learning methods to images, we must first extract a representation of each facial image as a feature vector, which encode some information about the patterns and/or structure in the face. Originally developed for texture analysis, Local Binary Pattern (LBP) features have since been instrumental in designing performant facial analysis algorithms. The original LBP descriptor thresholds a 3x3 neighborhood of each pixel with its center value and represents the result as an 8-bit binary number starting with the top left neighbor. A feature vector for a given image is then created by building a histogram using the LBP results over each pixel in the image [3]. Intuitively, one can think of each unique binary LBP result as encoding a “micro-pattern” in the face, including different types of curved edges, corners, line endings, etc. The full histogram contains information about the distribution of these patterns in the face. However, this feature representation does not represent the spatial location of these patterns, only their distributions over the entire image. To take into account the pattern spatiality, one can equally divide each image into a fixed number of small blocks, compute the LBP histogram over each block, and concatenate the histograms of each block into a single feature vector. By computing the distribution of micro-patterns of each individual block rather than the whole image, we can capture local pattern information and global structure information of each face [1].

For better feature extraction and pattern encoding, we extended the LBP descriptor to allow an arbitrary number of pixels within a radius of a circular neighborhood using bilinear interpolation. Our implementation used an 8 pixel neighborhood with a radius of 2. Still, our feature vectors remain relatively large, with each region’s histogram containing 2^P feature values, where P is the number of pixels in the neighborhood [1]. Research from [3] has shown that there are certain uniform patterns that contain more information than others. A pattern is uniform if its binary representation contains at most two bitwise transitions between 0 and 1 or vice versa [3]. In order to reduce the size of our feature vectors, we mapped each non-uniform pattern to a single bin, giving us P(P+1) + 3 bins for each block in the image.

Local Phase Quantization

Like LBP, the Local Phase Quantization (LPQ) descriptor was originally developed for texture classification and has also been successful in face recognition and expression analysis. The LPQ method is based on the blur invariance property of the Fourier phase spectrum by extracting the local phase information from a short-term Fourier transform (STFT) computed over an M-by-M rectangular neighborhood ***Nx*** at each pixel ***x*** in the image ***f(x)***.

$$F(\mathbf{u},\mathbf{x}) = \sum_{\mathbf{y}\in N_x} f(\mathbf{x}\text{-}\mathbf{y})e^{-j2\pi\mathbf{u}^T\mathbf{y}} = \mathbf{w}_{\mathbf{u}}^T\mathbf{f}_{\mathbf{x}}$$

For each pixel, the local fourier transform coefficients are computed at four frequency points

$$\mathbf{u}_1 = [a,0]^T,\ \mathbf{u}_2 = [0,a]^T,\ \mathbf{u}_3 = [a,a]^T,\ \text{and}\ \mathbf{u}_4 = [a,-a]^T$$

resulting in the vector:

$$\mathbf{F}_{\mathbf{x}}^c = [F(\mathbf{u}_1,\mathbf{x}), F(\mathbf{u}_2,\mathbf{x}), F(\mathbf{u}_3,\mathbf{x}), F(\mathbf{u}_4,\mathbf{x})]$$

Then the signs of the real and imaginary components of each coefficient are then quantized using a simple quantizer given by

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0 \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

where $q_j(x)$ is computed over the real and imaginary components from the result of the STFT computed at each of the four frequencies. The resulting 8 bit binary number can then be mapped into a bin of a 256 bin histogram [31]. Like LBP, block processing can also be applied to LPQ in order to encode the global structures in the image.

LBP/LPQ on Three Orthogonal Planes (TOP)

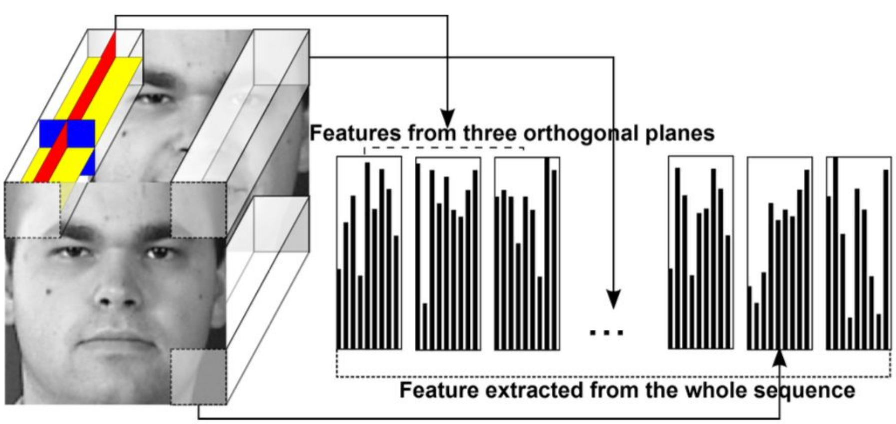

Figure 1. Features extracted from the XY, XT, and YT planes [32].

LBP and LPQ are static feature descriptors that operate on single images. However, when we examine image sequences, we can also exploit the information present in the temporal domain.

Traditionally, one thinks of an image sequence as a stack of XY planes concatenated along the time domain T. On the other hand, we can also think of an image sequence as a stack of XT planes concatenated along the Y domain or a stack of YT planes concatenated along the X domain. LBP/LPQ along Three Orthogonal Planes (TOP) extends LBP and LPQ to the time domain by extracting features along three orthogonal XY, XT, and YT planes as shown in figure 1. The resulting histogram computed over each plane is then concatenated to form the final feature descriptor of the image sequence. Note that we only take the three orthogonal planes depicted in the figure with the XY plane at the center of the sequence. We can obtain a better description if we extracted features from all the planes in the XY, XT, and YT stacks, however this can quickly lead to a very large feature space, which can cause overfitting during classifier training even with smaller dimensional image sequences [32].

Methods

First, we evaluate the effectiveness of LBP and LPQ feature descriptors in facial expression recognition systems. We then evaluate LBP-TOP and LPQ-TOP feature descriptors effectiveness in recognizing micro-expressions in image sequences. The current state of the art in micro-expression recognition employs LBP-TOP features and our contributions to the field include the evaluation of LPQ-TOP features. Our FER and MER system implementations both contain three parts: (a) preprocessing and image normalization, (b) facial representation and feature extraction, and (c) classifier training and validation.

Databases and Preprocessing

For our facial expression recognition system, we used the extended Cohn-Kanade (CK+) database [2], one of the most popular and comprehensive facial-expression image databases. All image sequences show subject's faces in full frontal view with little horizontal or vertical rotation. During the preprocessing stage, we selected all of the image sequences in the database for which there was an emotion label and picked the peak 3 image frames as well as the neutral frame for our training data, giving us a total of 1233 images in our training set [1]. For each image, we used the libraries in Matlab's Vision toolbox to find the bounding box of the left and right eyes, taking the center of the box to be the location of the pupil. We then rotated each image so that a line connecting the pupils would be parallel to the top boundary of the image. For normalization, we scaled each rotated image to a constant distance between the pupils. Finally, we cropped each image so the width measures double distance between the pupils and the height measures triple that distance, which were heuristics reported in [17].

Micro-expression recognition presents some additional and unique challenges over facial expression recognition. In the research community, several micro-expression databases have emerged, however there are few well-designed datasets which contain a sufficient quantity of elicited spontaneous micro-expression samples necessary for analysis. Some researchers created databases by asking a small number of participants to perform posed microexpressions very quickly, though this is very different from truly spontaneous expressions. Due to the short interval in which MEs occur, the image sequences captured with 20-30 fps off-the-shelf cameras do not contain enough ME frames to be useful in training a machine learning model.

A good micro-expression database should support the following requirements: (1) contain elicited spontaneous micro-expression sequences, (2) have a sufficient number of samples for analysis, and (3) contain higher spatial and temporal resolution samples. We chose the CASME II database because it supports these requirements. All sequences in the dataset were captured with 200 fps camera and all samples have a large face size of about 280 pixels by 340 pixels. In addition, the researchers elicited spontaneous micro-expressions from participants by asking them to suppress their facial expressions while viewing emotional videos. The researchers simulated a high-stakes situation by forcing participants to take a long and boring survey if they were caught displaying a facial expression. The ground truth in the database was established by two facial action coders who independently labeled each image sequence into one of five labels: happiness, disgust, surprise, repression, and others. Notice that these labels are slightly different than the ones present in standard expression recognition databases due to the difficulty of micro-expression elicitation and classification. Image sequences which were too short or contained expressions that were too subtle to be precisely coded were discarded. In total, the database contains 247 micro-expression samples from 26 participants. For our system, we took the cropped and normalized CASME II database as training data. We refer readers to [27] for details on the preprocessing algorithm.

Furthermore, we ran experiments where we applied Eulerian Motion Magnification (EMM) to each video sequence as a preprocessing step before feature extraction. The basic approach in EMM is to consider the time series of color values at any spatial location (pixel) and amplify variation in a given temporal frequency band of interest. Low-order IIR filters can be useful for both color amplification and motion magnification. We used the EMM which uses two low-pass IIR filters with cutoff frequencies ω_l and ω_h to create a bandpass filter to pull out the motion and signals that we wish to be amplified (facial expression changes in both color and geometry). For mathematical details on how EMM algorithm works, please refer to [26].

Feature Extraction

For our FER system, we extracted LBP and LPQ features from the CK+ database. For LBP, we obtained the best results with a neighborhood of size 8, a radius of 2, and a block size of 21 by 16 pixels. We also mapped non-uniform patterns to a single bin to reduce the size of the feature space, resulting in a 59 bin histogram for each block in the image. For LPQ, we obtained the best results with a neighborhood size of 5 and a block size of 35 by 25 pixels. For ME recognition, we extracted LBP-TOP and LPQ-TOP from the images sequences in the CASME II database. For LBP-TOP, we obtained the best results with a neighborhood of size 8 and a radius of 2 in each of the three orthogonal planes. For LPQ-TOP, we used a neighborhood of size 5 for the XY and XT planes and a neighborhood of size 15 in the YT plane.

Classifier Design

After creating feature vector representations for both FE recognition and ME recognition systems, we implemented multi-class classification of facial expressions by training Support Vector Machine (SVM) classifiers using a one-vs-rest scheme, generating one classifier for each expression class. To do this, we used Matlab's fitcecoc function to create a multi-class classifier using error correcting output codes (ECOC). We will not cover the inner workings of SVM or ECOC here. We refer readers to [28] and [29] for tutorials on these topics.

Results

Facial Expression Recognition

Figure 2. Best prediction accuracies on CK+ database

Feature Extraction Method	Best Prediction Accuracy
LPQ	96.43 %
LBP	94.28 %

Figure 3. LBP Confusion Matrix

	Neutral	Anger	Contempt	Disgust	Fear	Happy	Sadness	Surprise
Neutral	293	6	7	4	0	1	2	0
Anger	6	117	0	4	0	0	1	0
Contempt	6	0	41	0	0	0	0	0
Disgust	1	2	0	167	0	0	0	0
Fear	0	0	0	2	69	1	0	0
Happy	0	0	0	1	0	186	0	0
Sadness	4	2	0	1	0	0	75	0
Surprise	3	0	1	1	0	0	0	229

Figure 4. LPQ Confusion Matrix

	Neutral	Anger	Contempt	Disgust	Fear	Happy	Sadness	Surprise
Neutral	296	3	8	1	3	1	1	0
Anger	4	120	0	3	1	0	0	0
Contempt	4	0	42	0	0	0	0	1
Disgust	1	2	0	165	1	1	0	0
Fear	1	1	0	0	70	0	0	0
Happy	0	0	0	0	0	187	0	0
Sadness	4	1	0	0	0	0	77	0
Surprise	4	0	2	1	0	0	0	227

Our results on the CK+ database in the 7-class with neutral expression classification problem were very good and inline with the state of the art in [1]. From figure 2 and the confusion matrices in figure 3 and 4, we observe that LPQ performed slightly better than LPB. The face region of each image had an average resolution of 166 x 111 pixels so the blur invariance property of the LPQ was advantageous to preserving the pattern details in the image, giving it a slight performance gain over LBP. In the failure cases, both methods tended to confuse non-neutral expressions for neutral expressions and vice versa. While both methods are relatively computationally efficient compared with other appearance based feature extractors like Gabor wavelets, it is also important to note that the LBP benefits from feature space reduction by mapping uniform patterns to a single bin so that each extracted LBP histogram only contains 59 bins rather than the full 256 that LPQ has. This greatly reduces the runtime of model training when using LBP as compared to LPQ.

While we did obtain state of the art results on this database, there are various limitations to our approach. Histogram based feature descriptors depend heavily on preprocessing steps to normalize facial images. The CK+ database contains frontal face images which have very little horizontal or vertical rotation. While we have adopted this constraint to limit the amount of preprocessing steps, our approach will likely fail on facial images that appear "in-the-wild" or facial images that contain higher degrees of rotation. Another requirement of our system is that each facial region in the image be normalized so that it appears in relatively the same location across all the images in the dataset. For example, we would want the right eye to always appear in the same block in all the images, since we want our histogram-based descriptor to always extract patterns for the right eye region from the same block.

Micro-expression Recognition

Figure 5. Leave one out cross validation accuracy of micro-expression recognition experiments

Method	**LOOCV Accuracy**
LBP-TOP	51.12%
LPQ on peak frame	45.56%
LPQ-TOP	**52.13%**
LPQ-XTYT	47.98%
LPQ-TOP + Motion Mag	**53.73%**
State of the Art (LBP-TOP + Motion Mag + Temporal Interpolation Model) [30]	67.21%

The results we obtained for micro-expression recognition with LBP-TOP on the CASME II database are comparable with the results obtained in [30]. Our best results using LPQ-TOP show that it performs slightly better than LBP-TOP as a feature descriptor. We also applied the Eulerian motion magnification algorithm as a preprocessing step before extracting features and we found that this gave a slight performance boost to LPQ-TOP. In our experiments we had to consider the tradeoff between the number of extracted regions and the dimensionality of the feature space since LPQ will generate a feature vector of size 256 for each region. While this was not a major issue in the FER system, it becomes a problem when LPQ is extended to three orthogonal planes because the number of features is tripled which can cause overfitting when training machine learning models. In fact, we ran experiments which varied the number of blocks along each dimension for LPQ-TOP and we found that decreasing the number of blocks (and thus limiting the number of total features) actually increased our accuracy even though this encodes less global facial structure information. We also tried reducing the dimensionality of the feature space by ignoring one of the XY, YT, and XT planes completely during feature extraction though we ultimately found that considering all three planes yielded the best performance. These results give evidence that effective feature selection of LPQ-TOP features can yield better results and is a research direction that can be explored in the future.

The current state of the art results on the CASME II database uses LBP-TOP features, EMM, and a temporal interpolation model that normalizes all image sequences to have uniform number of frames in the time dimension [30]. This temporal normalization yields a significant increase in prediction accuracy as it allows better feature descriptions on the YT and XT planes.

Applying such a model before extracting LPQ-TOP features would likely also give better results and should be explored in the future.

Conclusion

In this project, we explored the relative merits of two histogram-based feature descriptors applied to facial expression recognition. We found that while the LPQ descriptor yielded slightly superior results compared to LBP, it produces a larger number of features which increases classifier training time. We then extended these descriptors to the LBP-TOP and LPQ-TOP variants in order to consider the time domain in image sequences with micro-expressions. We compared the performance of these two descriptors in recognizing micro-expressions and found that LPQ-TOP yields comparable results to LBP-TOP, though it performs slightly better. Micro-expression recognition is a relatively new and unexplored area in the field of computer vision and additional research and database creation is necessary in order to push the state-of-the-art forward and create commercial systems capable of real-time recognition with near-perfect accuracy.

References

[1] C. Shan, S. Gong, and P.W. McOwan, "Facial Expression Recognition based on Local Binary Patterns: A Comprehensive Study," *Image Vision Comput.*, vol. 27, no. 6, May 2009, pp. 803-816.

[2] T. Kanade, J. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2000.

[3] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distribution, Pattern Recognition 29 (1) (1996) 51–59.

[4] M. Valstar, I. Patras, 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, vol. 3, 2005, pp. 76–84.

[5] M. Valstar, M. Pantic, Fully automatic facial action unit detection and temporal analysis, in: IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2006, p. 149.

[6] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (12) (1999) 1357–1362.

[7] G. Donato, M. Bartlett, J. Hager, P. Ekman, T. Sejnowski, Classifying facial actions, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (10) (1999) 974–989.

[8] M. Pantic, I. Patras, Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences, IEEE Transactions on Systems, Man, and Cybernetics 36 (2) (2006) 433– 449.

[9] Y. Tian, Evaluation of face resolution for expression analysis, in: CVPR Workshop on Face Processing in Video, 2004.

[10] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distribution, Pattern Recognition 29 (1) (1996) 51–59.

[11] M. Pantic, L.J.M. Rothkrantz, Facial action recognition for facial expression analysis from static face images, IEEE Transactions on Systems, Man, and Cybernetics 34 (3) (2004) 1449–1461.

[12] M. Pantic, L. Rothkrantz, Expert system for automatic analysis of facial expression, Image and Vision Computing 18 (11) (2000) 881–905.

[13] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan, Recognizing facial expression: machine learning and application to spotaneous behavior, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[14] I. Cohen, N. Sebe, A. Garg, L. Chen, T.S. Huang, Facial expression recognition from video sequences: temporal and static modeling, Computer Vision and Image Understanding 91 (2003) 160–187.

[15] Z. Zhang, M.J. Lyons, M. Schuster, S. Akamatsu, Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron, in: IEEE International Conference on Automatic Face & Gesture Recognition (FG), 1998.

[16] C. Padgett, G. Cottrell, Representing face images for emotion classification, in: Advances in Neural Information Processing Systems (NIPS), 1997.

[17] Caifeng Shan, Shaogang Gong and P. W. McOwan, Robust facial expression recognition using local binary patterns, in: Image Processing, 2005. ICIP 2005. IEEE International Conference on, 2005, pp. II-370-3.

[18] S. Polikovsky, Y. Kameda, and Y. Ohta, “Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor,” in *ICDP*. IET, 2009, pp. 1–6.

[19] S. Polikovsky and Y. Kameda, “Facial micro-expression detection in high-speed video based on facial action coding system (FACS),” *IEICE Transactions on Information and Systems*, vol. 96, no. 1, pp. 81–92, 2013.

[20] Q. Wu, X. Shen, and X. Fu, “The machine knows what you are hiding: an automatic micro-expression recognition system,” *Affective Computing and Intelligent Interaction*, pp. 152–162, 2011.

[21] T.Pfister,X.Li,G.Zhao,andM.Pietikainen,“Recognising spontaneous facial micro-expressions,” in *ICCV*. IEEE, 2011, pp. 1449–1456.

[22] X.Li,T.Pfister,X.Huang,G.Zhao,andM.Pietika ̈inen,“Aspontaneous micro-expression database: Inducement, collection and baseline,” in *FG*.

[23] J. A. Ruiz-Hernandez and M. Pietikäinen, “Encoding local binary patterns using the re-parametrization of the second order gaussian jet,” in *FG*. IEEE, 2013, pp. 1–6.

[24] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu, “Micro-expression recognition using dynamic textures on tensor independent color space,” in *ICPR*. IEEE, 2014, pp. 4678–4683.

[25] S.-J. Wang, W.-J. Yan, G. Zhao, X. Fu, and C.-G. Zhou, “Micro- expression recognition using robust principal component analysis and local spatiotemporal directional features,” in *ECCV Workshop*, vol. 8925, 2014.

[26] Wu, H. Y., Rubinstein, M., Shih, E., Guttag, J. V., Durand, F., & Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world.

[27] Yan, W. J., Li, X., Wang, S. J., Zhao, G., Liu, Y. J., Chen, Y. H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, *9*(1), e86041.

[28] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, *2*(2), 121-167.

[29] Dietterich, T. G., & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, 263-286.

[30] Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2015). Reading Hidden Emotions: Spontaneous Micro-expression Spotting and Recognition. *arXiv preprint arXiv:1511.00423*.

[31] Ojansivu, V., & Heikkilä, J. (2008). Blur insensitive texture classification using local phase quantization. In *Image and signal processing* (pp. 236-243). Springer Berlin Heidelberg.

[32] Zhao, G., & Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*(6), 915-928.