2015 IEEE International Conference on Systems, Man, and Cybernetics

Micro-Facial Movement Detection Using Individualised Baselines and Histogram-Based Descriptors

Adrian K. Davison and Moi Hoon Yap
School of Computing, Mathematics and Digital Technology
Manchester Metropolitan University
Manchester, United Kingdom
Email: {A.Davison, M.Yap}@mmu.ac.uk

Cliff Lansley
Emotional Intelligence Academy
Walkden, United Kingdom
Email: Cliff@eiacademy.co.uk

***Abstract*—Detecting micro-facial movements in a video sequence is the first step in realising a system that can pick out rapid movements automatically as a person is being recorded. This paper proposes a new method of micro-movement detection by applying Histogram of Oriented Gradients as a feature descriptor on our in-house high-speed video dataset of spontaneous micro-facial movements. Firstly the algorithm aligns and crops faces for each video using automatic facial point detection and affine transformation. Then a de-noising algorithm is applied to each video before splitting them into blocks where the Histogram of Oriented Gradient features are calculated for each frame in every video block. The Chi-Squared distance measure is then used to calculate dissimilarity in the spatial appearance between frames at a set interval. The final feature vector is calculated after normalisation of the raw distance values and peak detection is applied to 'spot' micro-facial movements. An individualised baseline threshold is used to determine the value a peak must exceed to be classed as a movement. The result is compared with a benchmark algorithm - feature difference analysis techniques for micro-facial movements using Local Binary Patterns. Results indicate the proposed method achieves higher *Recall* of 0.8429 and *F1-measure* of 0.7672.**

***Index Terms*—Micro-Facial Movements, Facial Expression Analysis, Temporal Feature Extraction, Histogram difference, HOG**

I. Introduction

To detect the presence of a facial expression, videos are used to view the muscle movements of a person's face. In contrast, a person may look as if they are smiling on a static image but without a video to view the muscle motion, this cannot be confirmed. Static images cannot be used to spot micro-facial expressions, so video recordings must be used. The format of video used is also important as detecting subtle movements with the naked eye is much more difficult compared to the macro-facial expression counterpart [1].

Automatically recognising macro-facial expressions has become an established area in computer vision and machine learning [2], [3] often being able to work in real-time with relatively inexpensive equipment. Micro-facial expressions are becoming a more popular research topic due to the possibility of them being able to be used as an aid in detecting suppressed or repressed emotion [4], [5], [6]. However, detecting these expressions with computer vision is much more difficult than macro-expressions, as large facial expressions tend to show discriminative feature changes on the face, whereas micro-expressions are subtle, rapid changes that do not clearly fit into the discriminative 7 basic emotion classes - happy, sad, anger, fear, surprise, disgust or contempt [7].

A key to fully understanding micro-facial expressions is not to immediately associate them with emotions. The Facial Action Coding System (FACS) [8] objectively assigns Action Units (AUs) to the muscle movements of the face. If any classification of movements take place for micro-facial expressions, it should be done with AUs and not emotions. Emotion classification would require context of the situation so someone can make a meaningful interpretation. For the rest of this paper, the term 'micro-facial movements' will be used to describe the rapid and subtle expressions being detected to ensure objectivity. This paper proposes a new method for micro-movement detection by using Histogram of Oriented Gradients (HOG) as a feature descriptor to describe the video sequences. First, the algorithm will align and crop faces for each video in the chosen dataset using automatic facial point detection and affine transformation. A good feature descriptor is essential when modelling a micro-expression because if not appropriately chosen, the descriptor could lose important information or become overly sensitive to noise in the video recordings. As with any digital image capture, high-speed video can have large amounts of noise. To reduce this a de-noising algorithm is applied to the videos. The algorithm continues by splitting the videos into blocks and HOG features are calculated for each frame in every video block and the Chi-Squared distance measure is used to calculate dissimilarity in the spatial appearance between frames at a set interval. Finally, using a baseline that is unique to the participant in the dataset, a threshold value is calculated to highlight when a micro-movement is detected.

The rest of this paper is split into 4 sections. Section II describes the background and motivation of the research, and includes previous methods of micro-movement detection and recognition. Section III outlines the proposed method and Sec-

978-1-4799-8697-2/15 $31.00 © 2015 IEEE
DOI 10.1109/SMC.2015.326

tion IV details the micro-movement detection results compared with a benchmark method. Finally section V concludes the paper with future work and discussion.

II. Background

When a person consciously realises that a facial expression is occurring, the person may try to suppress the facial expression because showing the emotion may not be appropriate or a cultural display rule [9]. Once the suppression has occurred, the person will mask over the original facial expression and caused a transient facial change referred to as a micro-facial expression. In a high-stakes environment, micro-facial expressions become more pronounced [7], [5].

A. Micro-Expression Duration

The duration of micro-facial expressions are very short and this is considered the main feature that distinguishes them from a facial expression [10], with the general standard being a duration of no more than 500 ms [11]. Other definitions of speed that have been studied show micro-facial expressions to last less than 250 ms [4], less than 330 ms [12] and less than half a second [6]. Following Ekman and Friesen as first to define a micro-facial expression [1], a usual duration considered is less than 200 ms.

Experiments by Matsumoto and Hwang [13] summarise a micro-expression to be less than half a second with these experiments looking into whether training humans in detecting micro-facial expressions was effective. The findings showed that training improved the ability of reading micro-expression and the ability was retained a few weeks after the initial training. Training humans can be time consuming and expensive, so looking into ways of aiding a person when detecting subtle movements would make training more accessible.

B. Previous Work

There has been a limited work on detecting or recognising specific micro-facial expressions and movements to attempt to reduce the difficulty in human detection and limit the need for in-depth specialist training. The volatile nature of the appearance of these movements makes pattern recognition a challenge, and the lack of datasets compared with spontaneous macro-facial expression examples is also an issue. As this paper is focusing on spotting micro-facial movements in video rather than classifying particular emotions, the previous work will reflect this.

HOG features [14] was originally created for human detection in 2D images and used the pixel orientation values, weighted by its magnitude, to calculate features for describing a human as an object. Polikovsky et al. [15], [16] then extended this to a temporal descriptor that attempted to model micro-facial expressions. The recognition stage used *k*-means clustering to cluster particular AUs within the defined facial cube regions. The results were compared with ground truth ”tags” of muscle activation stages: neutral, onset, apex and offset. The classification rate for onset, apex and offset were 78.13% (80.02% with Transition Tags), 68.34% (70.99%) and 79.48% (81.85%) respectively. As a method for spotting micro-movements, it is able to show when a muscle was activated and then offset, however the dataset used for analysis contained posed micro-facial expressions and are not a good representation of naturally induced micro-facial expressions.

Wu et al. [17] use Gabor filters to automatically detect when recognising a micro-expression and spotting when they occur. However the training data used was 48 videos from METT [18] which flashes an expression very quickly to the user, but is classed as a posed expression like what is used in [15], [16]. These changes may be quick but frame by frame analysis should be able to pick up these unnatural movements easily.

Pfister et al. [19] use temporal interpolation with multiple kernel learning and Random Forest (RF) classifiers on their own spontaneous micro-expression corpus (SMIC dataset) [20]. The authors classify a MFE into positive or negative categories depending on two annotators labelling based on subjects' self reported emotions. Initial detection results on the 100 fps dataset was 74.3% and when down-sampled to 25 fps the accuracy increased to 78.9%. Like previous methods this required machine learning to classify if detection had occurred making the system unsuitable for real-time processing.

Using the gradient-based method of optical strain, Shreve et al. [21], [22] calculate the non-rigid facial motion of facial skin. A 78% true positive rate and .3% false positive rate was achieved for detecting micro-expressions. Further, this method is able to plot the strain and visualise a micro-expression occurring across time, however this can appear noisy due to head movements and talking that occur in the spontaneous micro-expression dataset used.

C. Benchmark Method

Moilanen et al. [23] follow an objective method that does not require any classification by using an appearance-based feature difference analysis method that incorporates Chi-Squared (χ^2) distance and peak detection to determine when a movement crosses a threshold and can be classed as a movement. After calculating the LBP histogram distance for videos split into 6x6 blocks, the top 1/3 blocks with the greatest difference values between frames are chosen to represent the initial feature vector of the face. A contrasting difference vector is then calculated to find relevant peaks from local magnitude variations and noise.

The threshold value for peak detection is set by taking the difference between the maximum and mean values of the contrasting difference vector and multiplying by a percentage value p that can be in the range of [0,1]. Finally this value is added to the contrasting vector mean to obtain the final threshold. By these calculations, the value will never exceed the maximum peak of the difference vector and thus not being able to detect neutral faces. The influence of the difference vector in threshold calculation causes at least one peak to always be classed as a detection. A baseline method is proposed in the next section to address this issue.

The datasets used are the CASME-A and B [24] and the original data from SMIC (not publicly accessible). For CASME-A the spotting accuracy (true positive rate) was 52% with 30 false positives (FP), CASME-B had 66% with 32 FP and SMIC-VIS-E achieved 71% with 23 FP. Only spatial appearance is used for descriptor calculation, therefore leaving out temporal planes associated with video volumes.

Previous methods mostly rely on classification in terms of emotion categories, which for micro-facial expressions is difficult to determine without context of the situation. The difference-based method proposed in Moilanen et al. [23] provides a more objective approach by using no machine learning or emotion categories, but by using a movement detection method. However, the approach uses a threshold determined by the movement sequence difference values, therefore the threshold will always detect at least one peak. In a real-world scenario, a more suitable approach would be to measure a person's baseline neutral expression and calculate a threshold based on this. If no movements occur, the peaks in the feature difference values will not exceed the baseline threshold.

In addition, previous datasets are limited in that they do not provide enough information to calculate a sufficient baseline in high-speed videos. An in-house dataset is used that consists of long high-speed video recordings of 200 fps where participants took part in an emotion inducement experiment.

III. Proposed Method

A. Face Alignment and Cropping

Normalisation is applied to all sequences so that all the faces are in the same position by using affine transformation. The points used for alignment are obtained using the Face++ automatic facial point detector [25]. To calculate the midpoint of the eyes, first the centre of both eyes are obtained automatically for both the left and the right eye and then the midpoint is computed as follows:

$$(M_x, M_y) = \left(\frac{LC_x + RC_x}{2}, \frac{LC_y + RC_y}{2} \right) \tag{1}$$

where M is the midpoint between the eyes and LC and RC are the centres of the left and right eye respectively. Using the calculated points, it can be worked out how to apply affine transformation to all images. First the distance between the eyes in Eq. 2 is found and then the angle between the eyes is calculated in Eq. 3:

$$(D_x, D_y) = (|RC_x - LC_x|, |RC_y - LC_y|) \tag{2}$$

$$\theta = \frac{\arctan(D_x, D_y)180}{\pi} \tag{3}$$

where D is the distance between the eyes and θ is the angle between the eyes. Using the extracted points this paper uses affine transform to align the eyes horizontally, ready to be processed. The face of the sequences then needs to be cropped to remove the unnecessary background in each image. Using this method does not require any manual annotation of points like in [23], and so is more suited to a fully automatic system.

Fig. 1. An example of a face divided into 5x5 blocks.

B. Dividing the Face into Blocks

The face is divided into blocks to preserve the local texture and global shape of the face [26]. This method has worked well for LBP with facial analysis tasks, and so it will be applied to HOG features to analyse its effectiveness. HOG originally performs some sliding window, block-based normalisation to obtain a final feature, however we leave this out to keep a standard across descriptors.

C. Difference Analysis

With all videos captured, some form of noise is certain due to the way images are captured digitally, whether this be through lighting, temperature or equipment malfunction. High-speed video is particularly susceptible due to capturing lots of images in a short space of time. To counter this, de-noising is applied using a sparse signal processing method [27] called collaborative filtering. This processes the video volume block-wise and attenuates noise to reveal fine details shared by the block groups while preserving the unique features of each block. The output after de-noising is the video of the same size so no further processing is required to suit the feature descriptors.

Spatial HOG features are extracted from each frame of each block of the video using Piotr Dollár's Matlab Toolbox [28]. This is done similarly to Dalal and Triggs [14] in that the pixel orientation and magnitude values are calculated and then binned into a histogram based on the chosen orientations. The original gradient computation was used where Gaussian $\sigma = 0$ (no smoothing) and the derivative mask used is a simple 1-D [-1,0,1] mask.

Originally the amount of orientation bins chosen in [14] are 9, but to model the movements of facial muscles in the spatial domain our method uses 8 bins. Also, unlike the original paper, the proposed method uses 2π for signed gradient orientation binning (see Fig. 3) so as to model the different directions the gradients can take and its relevance to facial expressions. Pixel magnitude (Eq. 4) and orientation (Eq. 5) are calculated and the magnitude values are binned into particular orientations so that the values are weighted based on the orientation of the

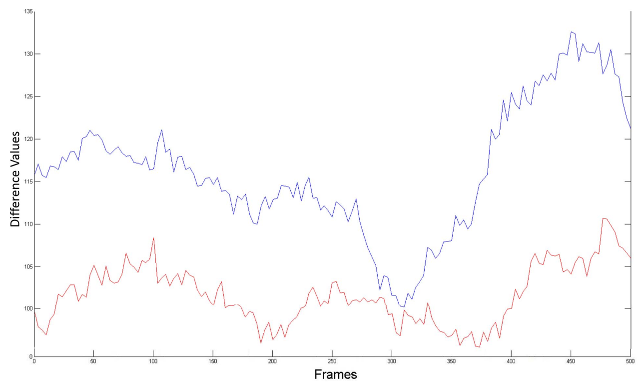

Fig. 2. The top blue line shows the higher values associated with a movement and the bottom red line shows the baseline feature values of a participant's neutral expression.

gradient.

$$Magnitude(x, y) = \sqrt{(\delta_x)^2 + (\delta_y)^2} \quad (4)$$

$$Orientation(x, y) = \arctan\left(\frac{(\delta_x)^2}{(\delta_y)^2}\right) \quad (5)$$

where δ_x and δ_y are the derivatives of the x and y spatial directions respectively.

The χ^2 distance [23] is then applied to obtain a feature vector of the de-noised HOG features of the in-house dataset. Instead of using the threshold value in [23] (henceforth *T1*) we propose an individualised baseline threshold (henceforth *T2*) obtained from the neutral videos of participants. The *T2* value is computed by taking a neutral video sequence for the participants and using the χ^2 distance to get an initial feature for the baseline sequence. The maximum value of this baseline feature is then used as the *T2* value. An example of the features extracted from a baseline video sequence and a movement video sequence is illustrated in Figure 2. The histograms produced from these features are able to discriminate between a movement and non-movement sequence as large deviations from the baseline indicate movement.

D. Performance Measure

To measure the performance of the algorithm, and to maintain the consistency with previous methods, it is important to outline what will be classed as a detection when a peak crosses a threshold. As all AUs have been FACS coded in the dataset, all micro-movements up to 100 frames that are detected will be classed as a true positive. Also, blinks and eye gaze are potentially useful when classifying micro-movements [4], [23] so these are classed as true positive. False positives are classed as peaks that occur but are not coded as a movement, blink or eye gaze. This mostly includes subtle head movements that can be noticed when using a high frame-rate camera.

We present our results in three measurements: *Recall* as in Eq.6, *Precision* as in Eq.7, and *F1-measure* (F1) as in Eq.8.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

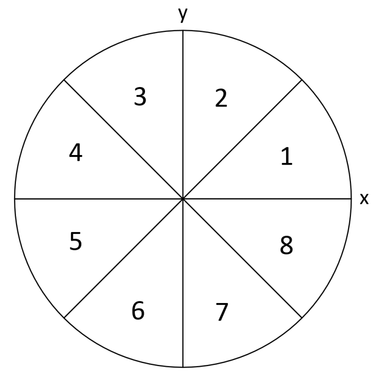

Fig. 3. Orientations of the 8 bins of the HOG descriptor.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \quad (8)$$

By using the *Precision* measure of exactness, the number of false positive detections can indicate if the method is over-sensitive and detects too many peaks. The F1 is useful in determining the harmonic mean between the *Precision* and *Recall* and is used in place of accuracy as it provides a more detailed analysis of the data.

IV. Experiments and Results

The dataset was obtained in an experiment comprising of 7 stimuli that each attempted to induce 1 of the 7 basic emotions in 32 participants. They were told to suppress their emotions and keep a neutral face. While actively trying to stop their facial expressions micro-facial movements may occur. To increase the chance of this happening, a prize of £50 was offered to person if they hide their emotion the best, therefore introducing a high-stakes situation [7], [5]. Each participant completed a questionnaire prior to the experiment so that the stimuli are tailored to each individual. The dataset was recorded at a higher temporal resolution of 200 fps and also a higher spatial resolution of 2040 x 1088 pixels. An example from the dataset can be seen in Figure 4. Overall, 29 participants exhibited micro-facial expressions and were FACS coded by 3 certified coders with an inter-coder reliability of 0.82.

A. Detection Results

The results shown in Table I outlines statistics on detection rates using the in-house dataset on the micro-movement detection using *T1* and the proposed method. Using *T2* calculated using the baseline of a participant's neutral facial expression, the results indicate a higher performance. The proposed method is able to spot a large number of micro-movements when using a higher temporal resolution of 200 fps and when using the spatial HOG descriptor compared with the LBP method.

Fig. 4. An example of a coded micro-movement. The movement is AU 4 and AU 7, where the brow is lowered and eyelid tightens. Image (a) is the onset frame, (c) is the apex where the brow is lower than the first image and some eyelid wrinkling is noticeable. Finally, (e) shows the participant returning to neutral.

The best result was *Recall* at 0.8428 using the proposed method and the *T2* threshold. This is much higher than the performance using LBP or HOG with the *T1* threshold with 0.5171 and 0.4657 respectively. Due to the dataset being high-speed videos and a higher resolution compared with previous experiments, it is likely that calculating distance will lead to large difference value peaks. *T1* was calculated using the mean and maximum values of the feature vector, so it will always detect a peak. This has the disadvantage in a real-world scenario if there are no movements (neutral face or no expression) they will always be misclassified as a movement. Using *T2*, the neutral expression for the participant is used as a threshold, therefore it is not affected by the values of the target sequence.

TABLE I
RESULTS USING *T1* AND *T2* INCLUDING *Recall*, *Precision* AND *F1-measure*.

Method	***Recall***	***Precision***	***F1-Measure***
LBP - [23]	0.5171	0.6084	0.5595
HOG - with *T1*	0.4657	0.7181	0.5650
LBP - with *T2*	0.7829	0.6508	0.7108
HOG - Proposed method	0.8429	0.7041	0.7672

Figure 5 shows how a large peak increases the *T1*, whereas the proposed method detects the large peak but also peaks that have been missed due to using the sequence in the threshold value calculation. In Figure 6 *T1* uses the neutral sequence to calculate a threshold that appears to have detected movements, but in fact are the baseline neutral expression of that particular participant. Further, when a neutral sequence is input, the baseline stays above the peaks as they are below the participant's baseline. Due to the proposed method's sensitivity, the *Precision* is affected and more false positives are observed. This can be seen in Table I as the *Precision* value between methods is relatively small, and using *T1* with HOG has a better *Precision* value of 0.7181 compared with using HOG and *T2* which achieved 0.7041. Further tuning of the baseline selection would be required to reduce false positives but preserve the sensitivity of micro-movements detection.

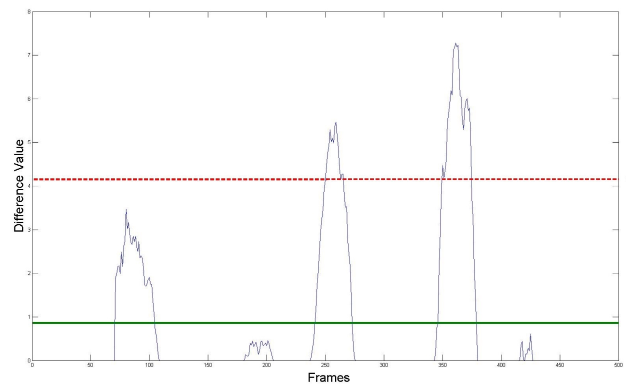

Fig. 5. Illustration of a micro-movement sequence where a micro movement (peak 1) is missed by *T1*. The green solid line shows *T2*, which detects all micro-movements.

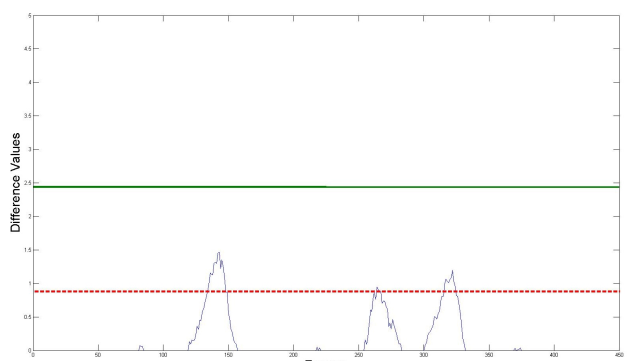

Fig. 6. The green solid line shows that *T2* stays above values that are considered to be a person's baseline. The dashed red line shows peaks falsely detected by the previous detection method using *T1*.

V. CONCLUSION

This paper has shown the potential of detecting micro-movements automatically using an individualised baseline based on a neutral video sequence and HOG features. The proposed objective method allows for analysis without assuming someone is attempting to hide a particular emotion. Finding movements without the assumption bias is crucial to ensure the most accurate findings can be found. For example, attempting to induce an emotion in a person may cause them to exhibit micro-expressions, however it is important to know why this person felt the need to suppress or repress their emotion before a category can be assigned to the movement.

Future work could also look into difference distance measures aside from Chi-squared. One such measure is the Earth Mover's Distance which can work well with normalised histograms to analyse the feature changes in the spatial, and also the temporal planes of a video.

Further, the method could be applied using different feature descriptors and more datasets. One area of interest would be to directly compare methods using high-speed videos with normal 30 fps videos to see what happens to detection rates with a high temporal resolution variance. Testing other datasets can also determine if the spatial resolution of images allows difference methods to perform better due to more information, or worse due to the information causing noise or more false positives. Extending the wrinkle detection method of a Hybrid Hessian Filter [29] will allow subtle edges of the face to be enhanced and applied to a temporal resolution to increase micro-movement detection rates.

More research into detecting micro-movements in real-time would be required so methods can be used in a real-world environment. Current methods require processing lots of computationally expensive feature descriptors that would be not realistic for high-speed and high resolution videos. Lower frame rates may allow quicker processing, but at the cost of a reduction in temporal information. When outside of a lab-environment, natural human behaviour such as talking, head movements and gestures can cause problems for the current method, which has not been tested on the dataset with such nature. Finding micro-movements hidden within these natural human actions would be the ideal case.

REFERENCES

[1] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception." *Psychiatry: Journal for the Study of Interpersonal Processes*, 1969.

[2] M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[3] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, pp. 259–275, 2003, cited by 947.

[4] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. Norton, 2001.

[5] P. Ekman, "Lie catching and microexpressions," in *The Philosophy of Deception*, C. W. Martin, Ed. Oxford University Press, 2009.

[6] M. Frank, C. Maccario, and V. Govindaraju, "Behavior and security," in *Protecting airline passengers in the age of terrorism*. Greenwood Pub. Group, 2009.

[7] P. Ekman, *Emotions Revealed: Understanding Faces and Feelings*. Phoenix, 2004.

[8] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978.

[9] D. Matsumoto, S. H. Yoo, and S. Nakagawa, "Culture, emotion regulation, and adjustment." *Journal of personality and social psychology*, vol. 94, no. 6, p. 925, 2008.

[10] X.-b. Shen, Q. Wu, and X.-l. Fu, "Effects of the duration of expressions on the recognition of microexpressions," *Journal of Zhejiang University SCIENCE B*, vol. 13, no. 3, pp. 221–230, 2012.

[11] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, "How fast are the leaked facial expressions: The duration of micro-expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, 2013.

[12] P. Ekman and E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, ser. Series in Affective Science. Oxford University Press, 2005.

[13] D. Matsumoto and H. S. Hwang, "Evidence for training the ability to read microexpressions of emotion," *Motivation and Emotion*, vol. 35, pp. 181–191, 2011.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, vol. 1. IEEE, 2005, pp. 886–893.

[15] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor," in *International Conference on Imaging for Crime Detection and Prevention*, 2009, pp. 16–21.

[16] S. Polikovsky, Y. Kameda, and O. Yuichi, "Facial micro-expression detection in hi-speed video based on facial action coding system (facs)," *IEICE transactions on information and systems*, vol. 96, no. 1, pp. 81–92, 2013.

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

[18] P. Ekman, "Mett," *Micro Expression Training Tool*, 2003.

[19] T. Pfister, X. Li, G. Zhao, and M. Pietikainen, "Recognising spontaneous facial micro-expressions," in *International Conference on Computer Vision (ICCV)*, 2011, pp. 1449–1456.

[20] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline." in *International Conference on automatic Face and Gesture Recognition*, 2013.

[21] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatio-temporal strain," in *International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 51–56.

[22] M. Shreve, J. Brizzi, S. Fefilatyev, T. Luguev, D. Goldgof, and S. Sarkar, "Automatic expression spotting in videos," *Image and Vision Computing*, vol. 32, no. 8, pp. 476 – 486, 2014.

[23] A. Moilanen, G. Zhao, and M. Pietikainen, "Spotting rapid facial movements from videos using appearance-based feature difference analysis," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, Aug 2014, pp. 1722–1727.

[24] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, "Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces," in *IEEE conference on automatic face and gesture recognition*, 2013.

[25] M. Inc., "Face++ research toolkit," www.faceplusplus.com, Dec. 2013.

[26] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, Dec 2006.

[27] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3d transform-domain collaborative filtering," in *Proc. 15th European Signal Processing Conference*, vol. 1, no. 2, 2007, p. 7.

[28] P. Dollár, "Piotr's Computer Vision Matlab Toolbox (PMT)," http://vision.ucsd.edu/ pdollar/toolbox/doc/index.html.

[29] C.-C. Ng, M. H. Yap, N. Costen, and B. Li, "Automatic wrinkle detection using hybrid hessian filter," in *Computer Vision–ACCV 2014*. Springer, 2015, pp. 609–622.