Spotting Rapid Facial Movements from Videos Using Appearance-Based Feature Difference Analysis

Conference Paper · August 2014

DOI: 10.1109/ICPR.2014.303

3 authors:

Antti J. Moilanen
Aalto University

Guoying Zhao
University of Oulu

Matti Pietikainen
University of Oulu

Spotting rapid facial movements from videos using appearance-based feature difference analysis

Antti Moilanen, Guoying Zhao, Matti Pietikäinen
Center for Machine Vision Research
Department of Computer Science and Engineering
University of Oulu, Finland

***Abstract*—Spotting micro-expressions is a primary step for continuous emotion recognition from videos. Spotting in this context refers to automatically finding the temporal locations of the face-related events from a video sequence. Rapid facial movements mainly include micro-expressions and eye blinks. However, the role of eye blinks in expressing emotions is still controversial, and often they are considered as micro-expressions as well. In this paper a simple method for automatically spotting rapid facial movements from videos is proposed. The method relies on analyzing differences in appearance-based features of sequential frames. In addition to finding the temporal locations, the system is able to provide spatial information about the movements in the face. Micro-expression spotting experiments are carried out on three datasets consisting only of spontaneous micro-expressions. Baseline micro-expression spotting results are provided for these three datasets including the publicly available CASME database. Also an example of spatial localization of the spotted rapid movements is presented.**

Keywords—Affective computing; gesture and behavior analysis; facial expression recognition

I. INTRODUCTION

Spotting facial movements from a video refers to the problem of automatically finding the temporal locations of the face-related events from a sequence of frames. Movements in the facial area are usually related to facial macro- and micro-expressions, blinking of eyes, changes in gaze direction, or talking. In case of micro-expressions, the events to be spotted are rapid and the facial movements are more subtle than with the ordinary, macro-expressions [3].

Emotion recognition based on micro-expressions is a challenging task that is gaining more attention because of its possible applications in e.g. behavioral analysis, business negotiations, forensic investigations, and security systems [5] [7] [22]. Currently, many of the micro-expression studies are focused on classification of facial expressions from well-segmented videos that are already divided into temporal sequences containing the micro-expressions [15]. However, considering real-life application domains such as human-machine interfaces and security systems, automatically spotting the micro-expressions is a primary step for the process. In this paper, the focus of interest is on spotting of rapid facial movements which mainly include micro-expressions and eye blinks. According to some psychological studies, eye blinks can also be considered as one type of micro-expressions, possibly revealing suppressed emotions [5]. In addition to automatically finding the temporal locations of the facial movements from a video sequence, it would be useful to determine the spatial location of the movements in the facial area as well. Spatial information could be used in further classification of the spotted micro-expressions or masking unrelated facial movements that cause instability in the analysis.

This work was supported by the Academy of Finland and Infotech Oulu.

In this paper, a simple method for automatically spotting rapid facial movements from videos is proposed. Besides spotting the frames around which the actions occur, the system is able to provide information about the spatial location of the spotted movements. The method relies on analyzing differences in appearance-based features of sequential video frames within a specified interval. Proposed method does not require training or pre-labeling of the videos, which makes it adaptive to unseen videos. In this paper, Local Binary Pattern (LBP) algorithm is used as a feature descriptor for the difference analysis. The proposed method has five main steps: (1) three stable facial points are tracked through the video; (2) each face image is divided into adaptive block structure based on two independent measures of human face; (3) LBP features are calculated for each block; (4) dissimilarity of features for each block of sequential frames within a defined time interval is calculated using the Chi-Squared distance; (5) resulting difference matrix is handled by: (i) obtaining difference values for each frame by averaging a number of the highest block difference values, (ii) contrasting relevant peaks by subtracting the average of the surrounding frames' difference values from each peak, and (iii) using thresholding and peak detection to spot rapid facial movements from the video. Spatial locations of the movements in the face are obtained by retrieving the blocks corresponding to the highest difference values.

Spontaneous and posed facial expressions differ in both which facial muscles are moved, and in the dynamics of the movements. Especially in case of micro-expressions, posed facial movements are more visible and more intense than the spontaneous ones. Also, micro-expressions are difficult to fake because they are involuntary and rapid [5]. Spontaneous expression data are evidently more challenging to classify than posed expression data [21]. In this paper, micro-expression spotting experiments are carried out on three datasets containing 273 spontaneous micro-expressions. Two of the datasets are from the publicly available CASME database [25], and one is extracted from the original data of the SMIC corpus [10]. Baseline micro-expression spotting results are provided for these three datasets. To the best knowledge of the authors, this micro-expression spotting experiment is the first to use datasets consisting only of spontaneous micro-expressions.

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.303

II. Background

A. Micro-expressions

Micro-expressions are brief and involuntary facial expressions that usually appear when people are trying to conceal their true feelings, or are unintentionally showing repressed emotions [3] [8]. Micro-expressions are distinguished from macro-expressions by their duration and the magnitude of the facial motions. Facial movements related to micro-expressions are often very subtle, and the duration of a micro-expression is traditionally considered to be between 1/25 to 1/5 of a second [3] [5]. However, in different micro-expression studies the duration varies anywhere between 1/25 to half of a second [7] [16] [18]. Recent experiments of Yan *et al*. [24] suggest that the total duration of a micro-expression (from its onset to the apex and back to offset) is generally within a half of a second, which is the guideline in this paper as well.

B. Micro-expression spotting

Automatic spotting of micro-expressions from videos is a relatively new research field, which can be seen from the small number of related studies and publicly available databases. Currently, few spontaneous micro-expression databases including full annotations are released that are suitable for micro-expression spotting. The spontaneous micro-expression database SMIC [10] consists of short videos including only the frames of micro-expressions from the onset to the offset, so it is not suitable for spotting as it is. Instead, the CASME datasets [25] provide mostly suitable data, but also there are some videos that are not very reasonable for micro-expression spotting – shortest video lasting only 0.2 seconds.

Besides the small number of available standard datasets, few studies have been published about micro-expression spotting. Micro-expressions were detected in [16] using 3D gradient histogram descriptor with high spotting accuracy reported, but on a very small dataset with faked micro-expressions. Gradient based approach was used also in [18] where optical strain was used as a descriptor for facial macro- and micro-expression spotting in longer videos. Spotting accuracy of 80 % with 38 % false positive rate was reported for faked micro-expressions. For small dataset of genuine micro-expressions, spotting accuracy of 52 % with 62 % false positive rate was achieved with a note that the spontaneous micro-expression videos included global head movement and talking which are often misclassified as micro-expressions. Gabor filters were used for automatic micro-expression recognition in [23], including the spotting step as well. In 48 videos dataset collected from METT [6], they achieved high spotting accuracy using over 1000 training examples. However, the high accuracy is expected, as the videos in the METT dataset are produced by placing a flash of a micro-expression in the middle of several seconds of showing a neutral face image, which should be straightforward to spot using a frame-by-frame based method.

III. Algorithm

A. Facial point tracking

Three facial points are detected in the first frame and tracked through the rest of the video. Inner eye corners and nasal spine point were selected for tracking since they are stable and not affected by facial expressions [21]. In this experiment, facial point detection in the first frame is done manually but it could be done using automatic facial landmark detection as well. Facial point tracking is done by implementing the basic Kanade-Lucas-Tomasi (KLT) algorithm [20] from Matlab's Computer Vision System Toolbox.

B. Face cropping and division into blocks

Coordinates of the tracked points are then used for image alignment and dividing the facial area into blocks in the spatial domain. The in-plane rotation is corrected by rotating the images according to the angle between the horizontal line and the line connecting the inner eye corners. Block division is done for preserving both the local texture and global shape information of the face [1]. However, global head movements could cause instability in the spotting analysis if the blocks were not attached to the image. In order to keep the contents of each block still, the block structure is fixed to a stable facial landmark point. Based on measurements from the tracking results of several videos from SMIC and CASME databases, it was noticed that there are some instability in the inner eye-corner point coordinates caused by blinking of the eyes, keeping the eyes closed for a longer time, or due to occluding items as glasses. Instead, the coordinates of the nasal spine are very stable, and the tracker is able to follow them during global head movements as well. Thus, the block structure is fixed according to the coordinates of the nasal spine point.

Optimal number of blocks was determined based on preliminary spotting experiments. In 6x6 block structure the blocks include conveniently all the critical parts of the face – yet keeping the number of blocks as low as possible to maintain computational efficiency. Size of the block structure is defined on the basis of two measures of the human face: horizontal distance between the inner eye corners and vertical distance between the nasal spine point and the line connecting the inner eye corners. Horizontal and vertical facial measurements are statistically independent, and the ratio between different measures varies between individuals depending on e.g. race, sex, and age [9]. This way, the algorithm is adaptable to faces of different sizes and shapes. In this experiment, the datasets are well-controlled, so that the distance between the face and the camera is maintained quite well through each video. Thus, averages of the facial measures are used to calculate a constant block size for each video. In case there are more variation in the distance between the face and the camera, the facial measures can be calculated frame-by-frame. Examples of block structures are presented in Fig. 1.

C. Feature extraction with LBP

Since its original publishing in the mid 1990's [12], the Local Binary Pattern (LBP) approach has been successfully applied in various fields of computer vision e.g. texture analysis, object detection, and human actions recognition. Furthermore, LBP and its variants have been applied in face recognition and micro-expression classification [1] [15].

LBP procedure is as follows. For each pixel in a frame, it is compared to each of its P neighbor pixels along a circle. Bilinear interpolation is used when the sampling point does

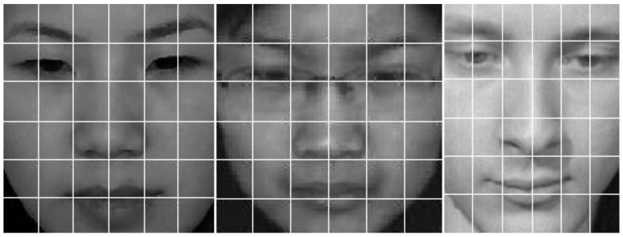


Fig. 1. Examples of block structures.

not fall in the center of a pixel. If the gray value of a neighboring pixel is greater than or equal to that of the center pixel, the output is 1; otherwise it is 0. This gives a P-digit binary number that can be presented as a decimal for simplicity by weighing with the powers of two. Thus, the LBP operator produces 2^P different output values [13]. Based on the LBP output matrix, a 2^P-bin histogram is computed. The images are divided into smaller blocks as explained in Section III-B, and LBP histogram is calculated for each block. After normalizing the block histograms they are concatenated to get one feature vector to represent the whole frame [1]. To reduce the length of the feature vector, uniform mapping is used [13].

D. Feature difference analysis

First, it is reasonable to denote some terms which are used in the following description. Current frame (CF) is, by its name, the frame that is currently analyzed. When a micro-interval of N (an odd number) frames is used, tail frame (TF) is the k-th frame before the CF, and head frame (HF) is the k-th frame after the CF, respectively, where

$$k = \tfrac{1}{2}\left(N-1\right). \tag{1}$$

Finally, average feature frame (AFF) is a feature vector representing the average of the features of TF and HF. It is temporally located at the same spot than the CF in a video sequence.

The basic idea of feature difference analysis is following: for each CF its features are compared to the respective AFF by calculating the dissimilarity of the feature vectors. The difference between the CF features and the AFF indicates the level of changes in the facial area. Moreover, the possible change in the features is rapid since it occurs between TF and HF, which distinguishes rapid facial movements from temporally longer events. This is repeated for each frame excluding the first k frames from the beginning and the end of the video, where TF and HF would exceed the video boundaries, respectively.

Let us take two occasions as examples. In the first case, a person in the video sequence starts to smile between TF and CF so that the smile is maintained also between CF and HF. This could be an onset phase of a macro-expression. In the second case, the person starts to smile between TF and CF, but the smile is quickly repressed between CF and HF. This would be a full micro-expression. Now, in the first case the calculated AFF would be more similar to the features of CF, than in the second case. In other words, the measured dissimilarity between the CF features and the AFF would be higher in the second case – especially when CF happens to be the apex frame of the micro expression. Similarly, other rapid facial movements such as eye blinks would yield to high dissimilarity measure between the CF features and the AFF, because the TF and HF features would be similar to each other, yet different from the CF features.

Dissimilarity of each pair of block histograms is calculated using the Chi-Squared (χ^2) distance. For normalized histograms P and Q with same number of bins, the χ^2 distance is defined as

$$\chi^2(P,Q) = \sum_i \frac{(P_i - Q_i)^2}{(P_i + Q_i)}, \tag{2}$$

where the index i refers to i-th bin in a histogram [14].

The χ^2 distance has been successfully applied in e.g. object and texture classification [26]. In many of the commonly used histogram distances the difference between large bin values is less significant than the difference between small bins [14]. However, in dissimilarity measurements the difference should not be dependent on the scale of the bin values when each feature is assumed to be equally important. Also Ahonen *et al.* [1] found that χ^2 distance performs better in face recognition than histogram intersection or log-likelihood distance.

E. Thresholding and peak detection

After obtaining the feature difference value for each block of each frame, the average of the M greatest block difference values is calculated for each frame in order to get an initial difference vector to represent the whole video. The initial difference vector F is hereby defined as

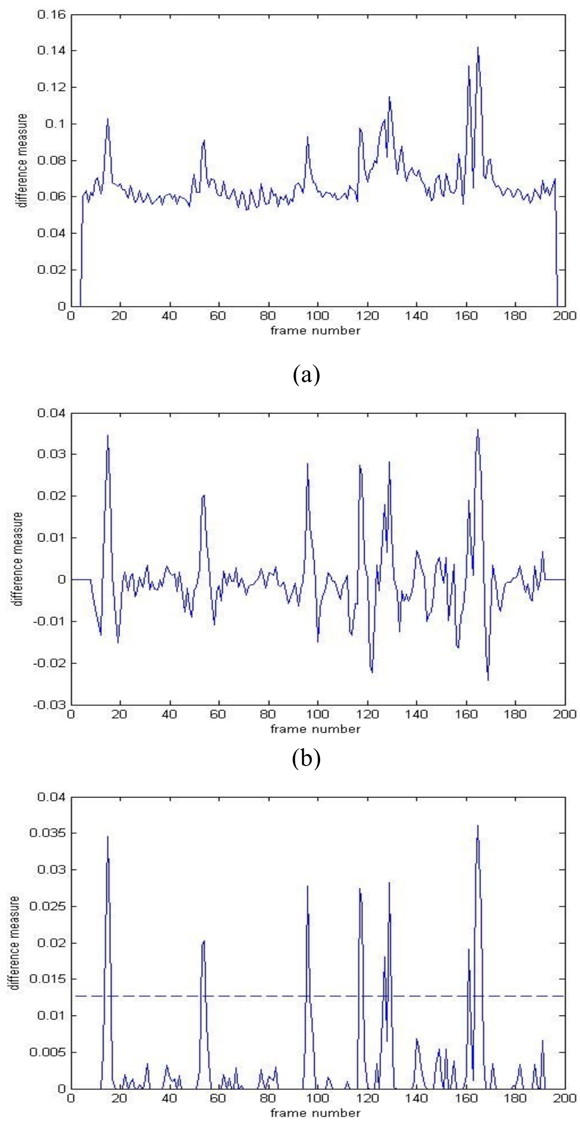
$$F = \frac{1}{M}\sum_{j=1}^{M}\left(D_{1,j}, D_{2,j}, \ldots, D_{n,j}\right), \tag{3}$$

where D is a (n x b) matrix containing the b block difference values sorted in a descending order for each frame, and n is the total number of frames. Using the 6x6 block structure, b is 36 in total. It was found that selecting about one third of the block difference values gives better discrimination than e.g. only one maximum value or average of all block values. Thus, M in (3) is set to 12. Example of an initial difference vector plotted with the difference values in the vertical axis and the frame numbers in the horizontal axis is presented in Fig. 2a.

To distinguish the relevant peaks from local magnitude variations and background noise, contrasting of the vector F is done by subtracting the average of the surrounding TF and HF initial difference values from each CF value. Thus, the i-th value in the contrasted difference vector is defined as

$$C_i = F_i - \tfrac{1}{2}\left(F_{i+k} + F_{i-k}\right), \tag{4}$$

and the contrasted difference vector for the whole video is obtained by calculating (4) for all n frames. However, it can be seen from (4) that contrasting cannot be done for the first and the last k frames of the video. The contrasted difference vector is presented in Fig. 2b. After contrasting, all negative difference vector values are assigned to be zero (see Fig. 2c). This is done because negative difference values indicate that the initial difference value of CF is below the average of initial difference values of TF and HF i.e. there are no rapid changes of



(a)

(b)

(c)

Fig. 2. Examples of a feature difference vector in different phases of the peak detection progress.

features in the CF compared to that of TF and HF. Finally, threshold and peak detection are applied to locate the peaks indicating the highest intensity frames of rapid facial movements. Threshold is calculated as

$$T = C_{mean} + p \times \left(C_{\max} - C_{mean}\right), \tag{5}$$

where C_{mean} and C_{max} are the average and the maximum of difference values for the whole video, and p is a percentage parameter in the range [0, 1]. Minimum peak distance in the peak detection is set to $k/2$. Threshold is illustrated in Fig. 2c as dashed line. In the example video of 200 frames, a micro-expression is successfully spotted around frame 127, and 7 blinks in the video are spotted in frames 15, 54, 96, 117, 129, 161, and 165. The example video is from the extracted SMIC-VIS dataset, which is discussed below.

IV. Experiments

A. Datasets and test setup

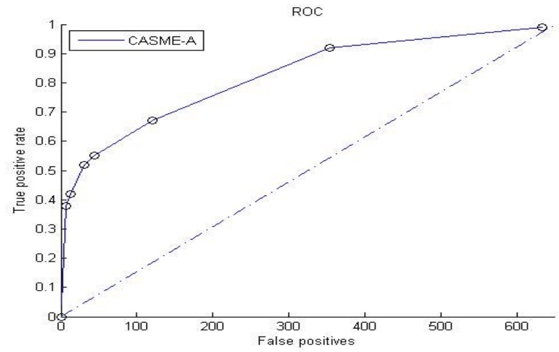
The Chinese Academy of Sciences Micro-expression (CASME) database [25] consists of 197 spontaneous facial micro-expressions from 19 subjects recorded with two cameras using rate of 60 frames per second. Ground truth annotation provides the frame numbers indicating the onset, apex, and offset frames. Total duration of micro-expressions in the database is less than 0.50 seconds or onset duration less than 0.25 seconds. The database includes two sections, which differ by lighting conditions and the camera used for recording. Frame resolution in CASME-A videos is 1280 x 720 pixels, and in CASME-B the resolution is set to 640 x 480 pixels. All samples from CASME database were used for evaluation. However, one video from CASME-B had to be handled with half the micro-interval than the others because of its short duration of only 13 frames. Moreover, in some of the CASME videos the micro-expression occurs right after the beginning or just before the end of the video. For fair evaluation, these samples were handled by determining that for the first and the last k frames of the video the first and the last frame are used as TF and HF, respectively. Average duration of videos in the CASME database is around 3.2 seconds. Minimum and maximum durations are 0.2 and 11.7 seconds, respectively.

The Spontaneous Micro-Expression Database (SMIC) consists of micro-expression video clips including only the frames from the onset to the offset [10]. Thus, it is not suitable for micro-expression spotting as it is. Instead, 71 longer videos were extracted from the original SMIC data and used for evaluation. The extracted videos were recorded using normal visual camera (VIS) with rate of 25 frames per second and resolution of 640 x 480 pixels. The released SMIC database includes samples recorded with a high speed camera as well, but due to problems in the video extraction they were not used in this experiment. Instead, the selected videos include all 71 micro-expressions from 8 subjects in the SMIC-VIS annotation and five additional micro-expressions found from the longer videos during the extraction. Average duration of videos is around 5.9 seconds. Minimum and maximum durations are 1.6 and 12 seconds, respectively. Below, this dataset will be referred to as SMIC-VIS-E as the extracted SMIC-VIS database.

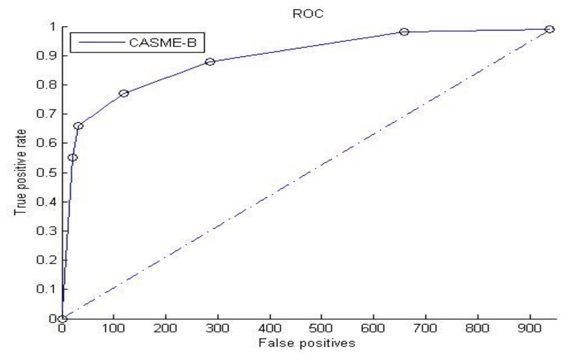
For CASME-A and CASME-B, a micro-interval of 21 frames was used, and for SMIC-VIS-E, the micro-interval was set to 9. Thus, according to (1) the parameter k is 10 and 4, respectively. Taking into account the frame rates, both correspond to about 0.16 seconds before and after CF. Spotted peak frames are compared with the provided ground truth frames, and considered true positive if they fall within the span of $k/2$ before or after onset or offset, respectively. Thus, in total the time span of true positive spots is 0.5 seconds, which is the presumed maximum duration of micro-expressions in this experiment. CASME datasets include some samples with no offset frame labeled. For those samples, the positive span was calculated by adding a number of frames according to the micro-interval to the labeled onset frame. In this experiment, spotted eye blinks are not counted as false positives. If needed, it is possible to use some classification method to distinguish the eye blinks from the micro-expressions after initial spotting.

B. Spotting results

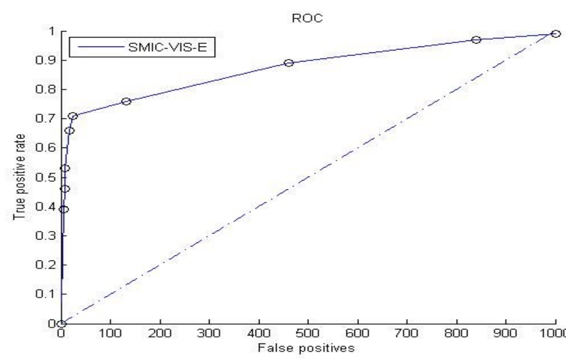
Spotting performance is evaluated using a Receiver Operating Characteristic (ROC) curve where – instead of the false positive rate – the number of false positives is presented in the horizontal axis as in facial feature detection [2]. True positive rate is in the vertical axis, and *p* in (5) is used as the discrimination threshold. ROC curves for the three datasets are presented in Fig. 3(a–c). Areas under the ROC curves (AUC) for CASME-A, CASME-B, and SMIC-VIS-E datasets are around 0.82, 0.90, and 0.90, respectively. Most of the points in the ROC curve are not reasonable due to high number of false positives. Instead, some examples can be given with a reasonable ratio between true and false positives. For SMIC-VIS-E dataset spotting accuracy of 71 % was achieved with 23 false positives using *p* of 0.30. For CASME-B dataset, 66 % of the micro-expressions were correctly spotted with 32 false positives using *p* of 0.85. For CASME-A dataset, true positive rate of 52 % wash achieved with 30 false positives using *p* of 0.65. Inferior spotting accuracy for CASME-A dataset may be attributed to higher video resolution, which could have required tuning of the parameters of the LBP descriptor for a better performance. For comparison, a spotting algorithm based on spatiotemporal strain [11] was applied in videos from the CASME datasets. However, the algorithm was extremely slow, and in a few videos that were tried, all of the micro-expressions were missed. In the future more comparative experiments are going to be conducted in order to see the performance of the proposed method in comparison with other methods.



(a)



(b)



(c)

Fig. 3. Micro-expression spotting performance for the three datasets.

C. Spatial information

Aside with temporal location, the feature difference analysis provides information about spatial location of the spotted facial movement as well. Example of block difference values over a face image from CASME-B dataset is presented in Fig. 4. The example is a successfully spotted apex frame of a micro-expression labeled as sadness consisting of AU:s 1+4+15 in the Facial Action Coding System (FACS) [4] i.e. inner brow raiser, brow lowerer, and lip corner depressor. All of the involved movements are evidently present among the 12 largest block difference values colored as white in Fig. 4.

D. Role of the eye blinks

According to neurophysiological studies, eye blinks can be classified into three types: reflexive, voluntary, and spontaneous. From these, spontaneous eye blinks are presumably related to the psycho-physiological state of a person [19]. Moreover, psychological studies suggest that blinking of eyes can be involved with several types of spontaneous muscular movements (with various meanings) e.g. squinting of eyes (facial gesture of nervousness or disagreement), rolling of eyes (contempt, disdain), uncontrollable blinking of eyes (stress relief), brow lowering (anger, dislike), and brow lifting (sadness, surprise) [5] [17]. Majority of the eye blinks are indeed reflexive, emotionally neutral, while some of them could be associated with repressed emotions i.e. to be evident micro-expressions.

To avoid unwanted spotting results caused by eye-related events, the eyes are usually masked [18]. However, during this experiment it was noted that masking the eye regions does not always prevent the eye blinks to be spotted. Instead, inspection of the block difference values indicated that blinking of eyes may cause rapid movements in several regions aside with the eyes e.g. in the eye brows and in the skin around eyes. Together with current knowledge about the kinematics and intention of eye blinks, this gives rise to the idea that it could be possible to distinguish the "meaningful" eye blinks from the reflexive, neutral eye blinks by using facial representation.

V. CONCLUSIONS

In this paper, a simple method for automatic spotting of rapid facial movements from videos was proposed. The method relies on analyzing differences in appearance-based features of sequential video frames, and does not require training or pre-labeling of the videos. This, together with dividing the facial area into a block structure defined by two independent measures of the human face and fixed according to a stable

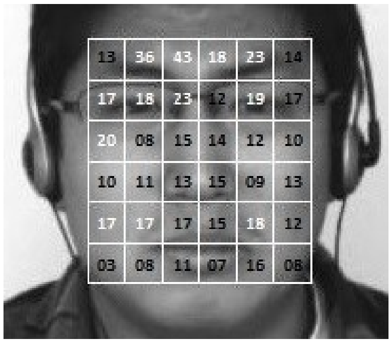


Fig. 4. Example of block difference values on a micro-expression frame.

facial point, makes the proposed method adaptive to unseen faces and videos. Features for the difference analysis are calculated using the Local Binary Pattern (LBP) algorithm. Experiments were carried out on three spontaneous micro-expression datasets with promising results. To the best knowledge of the authors, this is the first micro-expression spotting experiment that is done on datasets consisting completely of spontaneous micro-expressions.

In addition to automatically finding the temporal locations of the rapid facial movements, the proposed method is able to provide information about the spatial locations of the movements in the facial region as well. Spatial information can be used in further classification of the micro-expressions or masking unrelated facial movements that cause noise in the spotting analysis. During the experiment it was noticed that eye blinks can cause rapid movements in several regions aside with the eyes. According to psychological studies, some of the eye blinks might reveal suppressed emotions similarly than micro-expressions in general. Thus, in this experiment spotted eye blinks were not counted as false positives. Preliminary investigation of the block difference values in the spotted eye-related events indicated that it could be possible to distinguish the meaningful eye blinks from the neutral ones by using facial representation. In any case, further study and experiments on this topic are going to be conducted.

In the future, more adaptive threshold calculation is going to be developed for peak detection in order to improve robustness especially in longer videos. One possible solution would be to define local thresholds for shorter video segments, which could also enable real-time micro-expression spotting in the future. Large and rapid global head movements still cause instability in the analysis. In principle, the proposed method is able to spot rapid movements even from a side-view of the face as long as the facial movements are visible to the camera. However, if there are rapid changes in the pose, the system is likely to produce false positive spots. One solution for pose correction could be face image registration using e.g. affine transformation. Also, more comparative experiment is going to be conducted with some recently developed feature descriptors.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," in IEEE TPAMI, vol. 28, no. 12, pp. 2037–2041, 2006.

[2] M. Castrillón-Santana, O. Déniz-Suárez, L. Antón-Canalís, J. Lorenzo-Navarro, "Face and facial feature detection evaluation," in VISAPP, 2008.

[3] P. Ekman and W. Friesen, "Nonverbal leakage and clues to deception," in Psychiatry, 32(1), pp. 88–106, 1969.

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

[5] P. Ekman, Telling Lies – Clues to Deceit in the Marketplace, Politics, and Marriage. USA: Norton, 1992.

[6] P. Ekman, Micro-expression training tool, USA: University of California, 2003.

[7] P. Ekman, Lie Catching and Microexpressions. The Philosophy of Deception, ed. C. Martin, UK: Oxford University Press, 2009.

[8] E. Haggard and K. Isaacs, "Micro-momentary facial expressions as indicators of ego mechanisms in psychotherapy," in Methods of Research in Psychotherapy, USA: Appleton-Century-Crofts, pp. 154–165, 1966.

[9] M. Katsikitis (ed.), The Human Face: Measurement and Meaning. Kluwer Academic Publishers, 2003.

[10] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A Spontaneous Micro-expression Database: Inducement, collection and baseline," in FG, 2013.

[11] W. Miller, "Expression detection in video," Technical Report Comp Sci 534, USA: University of Wisconsin, 2011.

[12] T. Ojala, M. Pietikäinen, and D. Harwood , "A comparative study of texture measures with classification based on feature distributions," in Pattern Recognition, vol. 29, no. 1, pp. 51–59, 1996.

[13] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns," in IEEE TPAMI, vol. 24, no. 7, pp. 971–987, 2002.

[14] O. Pele and M. Werman, "The quadratic-chi histogram distance family," in ECCV, 2010.

[15] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in ICCV, 2011.

[16] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expression detection in hi-speed video based on Facial Action Coding System (FACS)," in IEICE Trans. on Inf. and Systems, vol. E96-D, no. 1, 2013.

[17] S. Porter and L. Brinke, "Reading between the lies: identifying concealed and falsified emotions in universal facial expressions," in Psychological Science, 19, pp. 508–514, 2008.

[18] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatiotemporal strain," in FG, 2011.

[19] R. Singla, B. Chambayil, A. Khosla, and J. Santosh, "Comparison of SVM and ANN for classification of eye events in EEG," in Journal of Biomedical Science and Engineering, 4, pp. 62–69, 2011.

[20] C. Tomasi and T. Kanade, "Detection and tracking of point features," Technical Report CMU-CS-91-132, USA: Carnegie Mellon Univ., 1991.

[21] M. F. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," in IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 42, no. 1, 2012.

[22] S. Weinberger, "Airport security: intent to deceive?," in Nature 465(7297), pp. 412–415, 2010.

[23] Q. Wu, X. Shen, and X. Fu, "The machine knows what you are hiding: an automatic micro-expression recognition system," in ACII, 2011.

[24] 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," in Journal of Nonverbal Behavior 37, pp. 217–230, 2013.

[25] 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 FG, 2013.

[26] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," in International Journal of Computer Vision, 73(2), pp. 213–238, 2007.