Quantifying Micro-expressions with Constraint Local Model and Local Binary Pattern

Wen-Jing Yan[1,2], Su-Jing Wang[1], Yu-Hsin Chen[1], Guoying Zhao[3], and Xiaolan Fu[1(✉)]

[1] State Key Lab of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
{yanwj,wangsujing,chenyouxin,fuxl}@psych.ac.cn, yanwj@wzu.edu.cn

[2] College of Teacher Education, Wenzhou University, Wenzhou, China

[3] Center for Machine Vision Research, University of Oulu, Oulu, Finland
gyzhao@ee.oulu.fi

Abstract. Micro-expression may reveal genuine emotions that people try to conceal. However, it's difficult to measure it. We selected two feature extraction methods to analyze micro-expressions by assessing the dynamic information. The Constraint Local Model (CLM) algorithm is employed to detect faces and track feature points. Based on these points, the ROIs (Regions of Interest) on the face are drawn for further analysis. In addition, Local Binary Pattern (LBP) algorithm is employed to extract texture information from the ROIs and measure the differences between frames. The results from the proposed methods are compared with manual coding. These two proposed methods show good performance, with sensitivity and reliability. This is a pilot study on quantifying micro-expression movement for psychological research purpose. These methods would assist behavior researchers in measuring facial movements on various facets and at a deeper level.

Keywords: Quantification · Micro-expression · Dynamic information · Constraint Local Model (CLM) · Local Binary Pattern (LBP)

1 Introduction

Micro-expression is usually defined as a brief and subtle facial movement revealing an emotion that a person tries to conceal [32]. Such characteristics make it a potential cue for lie detection [28]. It was claimed that well-trained inspectors reached 80% accuracy in lie detection based on micro-expression [12], which seems to be much more effective than other nonverbal cues. Micro-expression possesses theoretical implications and have many practical applications [13][32], but very few scientific researches were conducted on its characteristics.

A key reason for the lack of research on micro-expression may be due to it's difficult analysis [13]. Up till now, FACS remains to be the most widely used method for analyzing facial movements and many recent works (such as [25][27]

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part I, LNCS 8925, pp. 296–305, 2015.
DOI: 10.1007/978-3-319-16178-5_20

continue to use FACS to quantify behaviors. However, when trying to apply FACS to manually analyzing micro-expressions, researchers may encounter the following problems.

First, micro-expression is featured not only by its short duration [9] but also low intensity [24][32][31], some of which doesn't even reach the lowest intensity (level A) stated by FACS. These two characteristics usually make micro-expressions imperceptible to the naked eyes. Even with frame-by-frame approach, FACS coders find it difficult to spot the onset frame, apex frame and offset frame for many micro-expressions. As for describing dynamic information, it is impractical to describe increasing intensity over time via manual coding. However, research has shown the importance of dynamic information in facial expression [16][1][15][18]. Further quantification of dynamic micro-expressions may lead to interest findings; however, researchers lack suitable tools to quantify them. Second, manual coding with FACS is "arduous", especially for subtle facial movements [5]. As for very subtle facial expressions (a facial expression with intensity lower than the lowest intensity level according to FACS), coding would be even more difficult. Third, the coders, especially from different research groups, may follow a slightly different coding criterion even when the same coding system is employed (e.g. FACS). For example, when coding the onset frame of a facial expression, in Yan et al.'s [32] and Porter et al.'s [24] studies, the coders considered the onset frame as the first frame in which a change has occurred from the baseline (also see [14]). In FACS investigator's guide [10], however, Ekman defined "the first frame (film) or field (video) when the AU was at all visible" as the onset. This definition is vague and heavily dependent upon the coders' subjective judgment. With different research groups, the onset and offset frame coding may be difficult to replicate.

1.1 Utilization of Feature Extraction Methods to Facial Expression Analysis

Considering the issues rising from the use of FACS through manual coding, computer scientists have been trying to develop tools to analyze facial movements [4][7][11][17][20][26]. Researchers in the field of computer vision have previously focused on accurately classifying different facial expressions [3][22] and AUs [6][20][29]. However, for behavioral researchers, it would be more meaningful to know how the facial movements change in detail. Since psychological researchers still debate over the existence of universal categories of emotional facial displays, for behavioral science research purposes it would be more useful to quantify facial movements and further study the patterns of facial movements, rather than just providing a classification. In the following, we will introduce Constraint Local Model (CLM), which is mainly a geometric feature-based method, and Local Binary Pattern (LBP), which is an appearance-based method. CLM is improved from the commonly used Active Appearance Model (AAM) and Active Shape Model (ASM), and LBP has been applied to extracting features for micro-expression recognition [23]. These two algorithms seem suitable to deal with micro-expression.

1.2 The Aim of This Work

In this study, we select and apply two feature extraction methods to quantify dynamic information of micro-expressions, which would be facilitate psychological studies on micro-expression. CLM is employed to automatically detect the facial feature points. Based on the points, the faces were aligned and the ROIs (Regions of Interest) on the face were drawn for texture feature extraction by LBP. These two methods were evaluated by testing the effectiveness on quantifying micro-expressions. This paper provides a brief introduction to these algorithms, the way of applying them to quantifying dynamic information and test their performance on analyzing spontaneous micro-expressions.

2 Methods

2.1 Materials

Fifty micro-expression samples from CASME2 were selected. CASME2 is a spontaneous micro-expression database which contains 247 samples at 200fps (the inter-frame duration is 0.05s) and with the spatial resolution at about 280x340 pixels on facial area [30]. For demonstration, two samples (Fig. 1) were used to show how these two methods were applied to analyzing micro-expressions.

2.2 Applying CLM to Quantifying Micro-expressions

CLM CLM is a type of point distribution model (PDM), which represents the geometry mean and some statistical modes of geometric variation inferred from a training set of shapes [8]. CLM typically involves an exhaustive local search for the best location of each PDM landmark that is then constrained to adhere to the PDM's parameterization. We tested the source codes from [2], Jason Saragih's [1] and Yan Xiaoguang's [2], and found the Jason Saragih's tool performed best on CASME2. This algorithm trains 66 landmarks on the face. With the 66 landmarks detected on each frame, the coordinate of each landmark is generated. Since the facial movements may accompany some degree of head movements and even the slight head movement may blur or disrupt the targeted subtle facial movement, the coordinate of each landmark are subtracted by the coordinates of landmark 34, the nostril, which has a clear contour. Then we calculate the changes for each landmark. All video samples in CASME2 start with baseline (usually a neutral facial expression). To calculate the changes of each facial point, the corresponding coordinates of the 66 landmarks are subtracted by the pixel coordinates of the first frame. We get a graph depicting the changes of each landmark and form a "changing map" of the whole face, where the landmarks between 1 to 17 indicates the contour of the face, 18 to 27 indicates eye brows, and so on (see Figure 2).

[1] https://github.com/kylemcdonald/ofxFaceTracker

[2] https://sites.google.com/site/xgyanhome/home/projects/clm-implementation

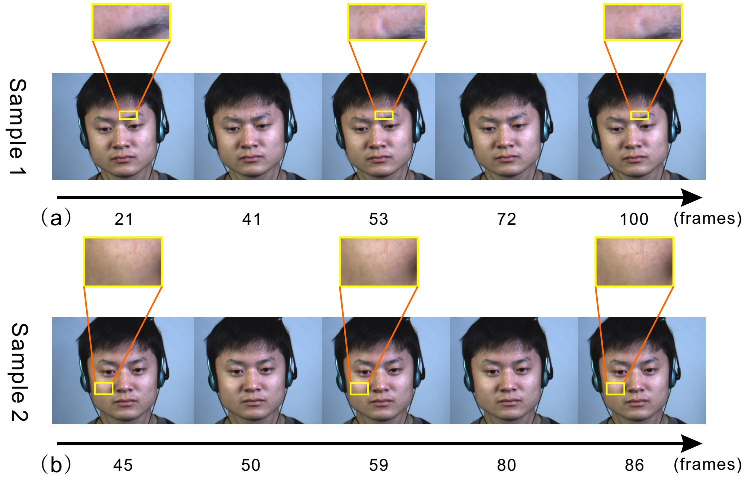

Fig. 1. The image-sequences of two samples from CASME2 that indicate the facial movements across time. (a) Zoomed-in images of the area of the left inner brow are shown in order to illustrate the movements. Frame 53 is the apex of the facial movement, while the frame 21 and frame 100 were taken as the baseline. The change is obvious for naked eyes. (b) Zoomed-in images of the area of the right cheek are to illustrate the movements. The apex frames are in 54-68 (too subtle to define), while frame 45 and 86 were taken as the baseline. The change is very subtle to detect, but better perceivable in video mode. Note: These samples were recorded by a 200fps camera, with an inter-frame duration 0.05s.

2.3 Applying LBP to Quantifying Micro-expressions

LBP For psychological research purpose, texture information of a certain region may be used to measure the change of the facial movements across time. Previous studies have shown that LBP is a powerful algorithm for texture description [21]. For a pixel C in the image, an LBP operator describes its local texture pattern by comparing and thresholding the gray values of its neighboring pixels against the gray value of pixel C. For the center pixel C with its P neighboring pixels sampled with the radius R, the LBP value is then calculated.

The source code are available here [3]. The following steps were carried out to quantify the aforementioned two samples (shown in Figure 1) of subtle facial movements with LBP. *Step 1.* Draw and select the region of interests (ROIs). Based on the landmarks detected by CLM, we draw the regions of interest (ROIs) for each frame. These ROIs were defined (partly) according to AUs. These regions includes the inner eyebrows (AU1, AU4), outer eyebrows (AU2), nose root (AU9), lower eyelid (AU7), cheeks (AU6), mouth corner (AU12, AU14, AU 15) and the regions at the side of the nose (AU 10), the jaw (AU17) and so on. The to-be-analyzed regions are selected by naked eyes. *Step 2.* Extract LBP for the ROI in each frame. *Step 3.* Calculate the change between the first frame

[3] http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab

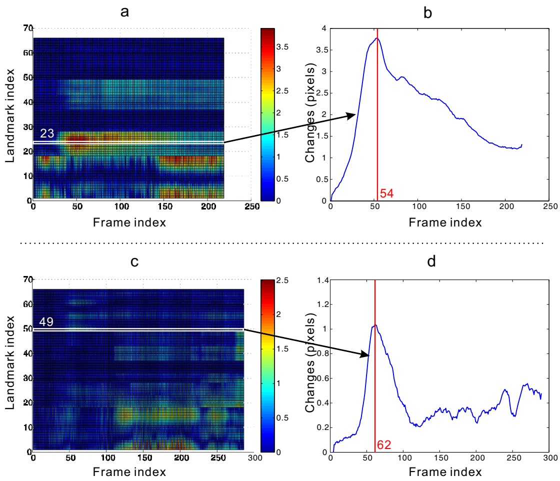

Fig. 2. The movements measured by calculating the changes of the landmarks across time. X-axis shows the frames of the video sample. (a) and (b) are for describing Sample 1, while (c) and (d) for Sample 2. In (a) and (c), y-axis indicates the numbered landmarks (from 1 to 66), and the color indicates the degree of change from the first frame. Landmarks 22 to 25 show great "activation" in (a), meaning that there is obvious change in the area of brows. The change of landmark 23 (left inner brow) is exhibited in (b), where y-axis means the location it changes.

and other frames for the ROIs. Similarity between two images is calculated by their correlation. And the rate of the texture change can be calculated by the difference between the first frame and the other frames. The correlation coefficient is calculated by:

$$d = \frac{\sum_{i=1}^{nBins} h_{1i} \times h_{2i}}{\sqrt{\sum_{i=1}^{nBins} h_{1i}^2 \times \sum_{i=1}^{nBins} h_{2i}^2}} \quad (1)$$

where h_1 indicates the gray-scale histogram of the first frame, h_2 the current frame. (1-d) indicates the rate of difference of the texture features in a ROI between these two frames. The peak of difference is found at frame 60 for sample 1, frame 59 for sample 2. The difference between the peak and first frame at left inner brow (sample 1) is about 0.018% and at right cheek is about 0.0025%. These values indicate the change of texture feature, which may serve as a measurement of intensity (Figure 3).

2.4 Data Analysis and Results

In the following, we compare the performance between the computer and manual coding in spotting the apex frames. It needs to be noted that the data obtained via manual coding don't necessarily represent the standard or correct answer as different people have different judgments.

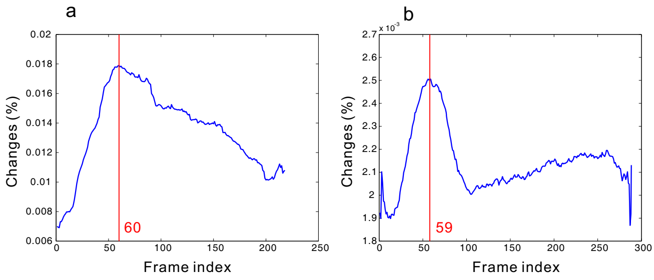

Fig. 3. The differences between the first frame and following frames across time for (a) Sample 1 (AU4, inner corner) and (b) Sample 2 (AU6, cheek)

CLM and LBP algorithms were tested on 50 samples of micro-expressions. Two coders coded the apex frame for each sample and the mean number of the two coded peaks was taken as the apex frame. To evaluate the performance of these two algorithms, the coded frame numbers by the proposed methods (CLM and LBP) and the manually coded on apex frame are compared (by subtracting the manually coded number). The result for CLM is: M=1.02, SE= 1.56; for LBP (subtracting manually coded number): M=0.31, SE=1.88. The difference (measured by frames) between the proposed methods and manual coding for each sample is demonstrated in Figure 4.

Analysis of variance (ANOVA) repeated-measures was conducted to test whether three coding methods, CLM, LBP and manual coding, are statistically different from each other (or whether from different population). Results show that the difference among these three methods is not significant, F (2, 76) = 0.273, p = 0.762, which indicates that the performance of the proposed algorithms matches those of manual coding in terms of spotting apex frames.

The details about CLM and LBP coding for different areas are demonstrated in Table 1. CLM and LBP have their own advantages when quantifying different areas of the face. The CLM seems to be well suited for measuring the areas around the eyebrow. LBP seems to outperform CLM in quantifying the area around the mouth, where the movement such as pressing lips have obvious texture change but not necessarily shape change.

3 Discussion

In non-verbal behavior studies, manual coding usually have suffered from subjectivity and inaccurate quantification. To avoid the use of heuristic coding scheme in measuring micro-expressions, this paper introduced two feature extraction methods, CLM and LBP, to quantify dynamic facial movements. CLM detects the 66 landmarks on the face for each frame and the detected landmarks are also used for alignment (which can partly deal with head motion problem). The texture features of the aligned ROIs (which are based on the landmarks detected by CLM) were extracted by LBP. These methods could be adopted to analyze

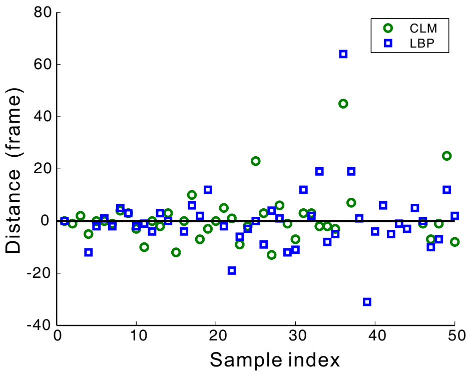

Fig. 4. The difference (in frames) between our proposed methods and manual coding on apex frame for each sample of micro-expression. The x-axis indicates sample No., and the y-axis indicates how many frames the proposed methods are off in comparison with the manual coding. The black line (y=0) represents the baseline (manual coding) and the blue squares and green circles represent the results of LBP and CLM respectively. Note: one frame is equivalent to 0.05 second in duration.

Table 1. The descriptive statistics for coding different areas with CLM and LBP

		CLM			LBP		
Area	Quantity	Reasonable	Mean	SE	Reasonable	Mean	SE
Eyebrow	16	16	3.31	2.99	14	3.00	4.92
Mouth	16	13	1.08	2.25	16	-4.38	2.85
Mix	5	3	5.00	10.1	5	8.00	3.02
Others	13	9	-4.22	1.89	12	0.08	1.38

micro-expression at a deeper level and reduce the amount of manual coding work. The following issues are some of the problems worth further discussion.

3.1 Sensitivity to Subtle Facial Movements

The values of the proposed methods for quantifying micro-expression mainly lies in its sensitivity and accuracy in quantifying the dynamic information of micro-expressions. CLM and LBP extract facial features with different approaches and each have their own advantages on different samples. CLM is sensitive to the eyebrow movement, which has a "clear" contour and thus CLM can accurately detect subtle movements near the eyebrow area. For LBP, the ROIs are initially drawn based on landmarks detected by CLM, the movement of the landmarks leads to the movement of ROIs (thus they are aligned). If the targeted areas don't show obvious texture change, the texture feature in these ROIs will not reveal changes as well. In contrast, for subtle movements such as a lip pressing, the landmarks of lip "corners" actually don't move too much, the gray-scale changes (texture changes) around the lip corner are obvious. The ROIs drawn based on by these landmarks are relatively steady, thus LBP is more suitable to measure these areas.

3.2 Measuring Dynamic (Temporal) Information

Dynamic information provides an abundance of information. Recently, researchers are paying more attention to dynamic facial expression instead of traditional static images of prototypical expressive patterns, since dynamic facial expressions are more ecologically valid [19]. With the advances of feature extraction methods, it is possible to measure and quantify the dynamic information such as moving distance, direction, velocity and texture feature, even on subtle facial movements. For CLM, the "changing map" of the facial movements (Figure 2) reveals the global characteristics which contain the information on the pattern of a facial expression. LBP doesn't show the direction but it can measure the global texture changes of the ROIs across time. These methods provide dynamic information that may reveal some specific patterns of facial expressions.

3.3 The Representative of Intensity

The strength of muscle contractions, acquired through electromyography, could be used to measure the intensity of a facial expression. For view-based analysis, however, the criteria for the intensity are difficult to define. In manual coding, coders define the intensity of a facial expression heuristically. For CLM, the intensity was calculated according to the displacement of the feature points across frames. For LBP, the intensity was represented by the texture (gray-scale pixels) change of a certain ROI across frames. LBP may be better since it considers information in a region instead of simply the coordinates of the landmarks.

3.4 Limitations and Future Work

Though feature extraction methods have promising applications in analyzing micro-expressions, there are still several challenges. First, landmark detection on the face is not always accurate and steady, even though these algorithms have made a remarkable progress over the past decade. Second, we have not established criteria for defining the onset and offset frame of micro-expressions. To achieve automatic detection of onset and offset frames for a given facial movement, future research should focus on the indicators and criteria for determining these key frames.

Acknowledgments. This work was supported by grants from 973 Program (2011CB302201), the National Natural Science Foundation of China (61379095, 61375009, 61175023, 61472138), China Postdoctoral Science Foundation funded project (2012M520428, 2014T70133), the Beijing Natural Science Foundation (4152055), and the Open Projects Program of National Laboratory of Pattern Recognition (201306295). G. Z. was supported by the Academy of Finland and Infotech Oulu.

References

1. Ambadar, Z., Cohn, J.F., Reed, L.I.: All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. Journal of Nonverbal Behavior **33**(1), 17–34 (2009)
2. Asthana, A., Zafeiriou, S., Cheng, S., Pantic, M.: Robust discriminative response map fitting with constrained local models. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3444–3451. IEEE (2013)
3. Bartlett, M., Littlewort, G., Whitehill, J., Vural, E., Wu, T., Lee, K., Eril, A., Cetin, M., Movellan, J.: Insights on spontaneous facial expressions from automatic expression measurement. Dynamic Faces: Insights from Experiments and Computation (2006)
4. Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J.: Measuring facial expressions by computer image analysis. Psychophysiology **36**(2), 253–263 (1999)
5. Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia **1**(6), 22–35 (2006)
6. Cohn, J.F., Sayette, M.A.: Spontaneous facial expression in a small group can be automatically measured: An initial demonstration. Behavior Research Methods **42**(4), 1079–1086 (2010)
7. Cohn, J.F., Zlochower, A.J., Lien, J., Kanade, T.: Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. Psychophysiology **36**(1), 35–43 (1999)
8. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models (2006)
9. Ekman, P., Friesen, W.: Nonverbal leakage and clues to deception. Tech. rep., DTIC Document (1969)
10. Ekman, P., Friesen, W., Hager, J.: Facs investigator guide. A human face (2002)
11. Essa, I.A., Pentland, A.P.: Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 757–763 (1997)
12. Frank, M., Ekman, P.: The ability to detect deceit generalizes across different types of high-stake lies. Journal of Personality and Social Psychology **72**(6), 1429 (1997)
13. Frank, M., Maccario, C., Govindaraju, V.: Behavior and security, pp. 86–106. Greenwood Pub. Group, Santa Barbara (2009)
14. Hess, U., Kleck, R.: Differentiating emotion elicited and deliberate emotional facial expressions. European Journal of Social Psychology **20**(5), 369–385 (2006)
15. Hess, U., Kleck, R.E.: The cues decoders use in attempting to differentiate emotionelicited and posed facial expressions. European Journal of Social Psychology **24**(3), 367–381 (1994)
16. Kappas, A., Descteaux, J.: Of butterflies and roaring thunder: nonverbal communication in interaction and regulation of emotion. In: Nonverbal Behavior In Clinical Settings, pp. 45–74 (2003)

17. Koelstra, S., Pantic, M., Patras, I.: A dynamic texture-based approach to recognition of facial actions and their temporal models. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(11), 1940–1954 (2010)
18. Krumhuber, E., Kappas, A.: Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. Journal of Nonverbal Behavior **29**(1), 3–24 (2005)
19. Krumhuber, E.G., Kappas, A., Manstead, A.S.: Effects of dynamic aspects of facial expressions: a review. Emotion Review **5**(1), 41–46 (2013)
20. Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M.: The computer expression recognition toolbox (cert). In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), pp. 298–305. IEEE (2011)
21. Ojala, T., Pietikinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition **29**(1), 51–59 (1996)
22. Pantic, M., Patras, I.: Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **36**(2), 433–449 (2006)
23. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Recognising spontaneous facial micro-expressions. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1449–1456. IEEE (2011)
24. Porter, S., ten Brinke, L.: Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. Psychological Science **19**(5), 508–514 (2008)
25. Rahu, M.A., Grap, M.J., Cohn, J.F., Munro, C.L., Lyon, D.E., Sessler, C.N.: Facial expression as an indicator of pain in critically ill intubated adults during endotracheal suctioning. American Journal of Critical Care **22**(5), 412–422 (2013)
26. Sebe, N., Lew, M.S., Sun, Y., Cohen, I., Gevers, T., Huang, T.S.: Authentic facial expression analysis. Image and Vision Computing **25**(12), 1856–1863 (2007)
27. Seidl, U., Lueken, U., Thomann, P.A., Kruse, A., Schrder, J.: Facial expression in Facial expression in alzheimers disease impact of cognitive deficits and neuropsychiatric symptoms disease impact of cognitive deficits and neuropsychiatric symptoms. American Journal of Alzheimer's Disease and Other Dementias **27**(2), 100–106 (2012)
28. Vrij, A.: Detecting lies and deceit: Pitfalls and opportunities. John Wiley & Sons Ltd., West Sussex (2008)
29. Wu, T., Butko, N.J., Ruvolo, P., Whitehill, J., Bartlett, M.S., Movellan, J.R.: Multilayer architectures for facial action unit recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics **42**(4), 1027–1038 (2012)
30. Yan, W.J., Li, X., Wang, S.J., Zhao, G., Liu, Y.J., Chen, Y.H., Fu, X.: Casme: An improved spontaneous micro-expression database and the baseline evaluation. PLoS ONE (2014)
31. Yan, W.J., Wang, S.J., Liu, Y.J., Wu, Q., Fu, X.: For micro-expression recognition: Database and suggestions. Neurocomputing (2014)
32. Yan, W.J., Wu, Q., Liang, J., Chen, Y.H., Fu, X.: How fast are the leaked facial expressions: The duration of micro-expressions. Journal of Nonverbal Behavior, 1–14 (2013)